

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
14 June 2001 (14.06.2001)

PCT

(10) International Publication Number
WO 01/42451 A2

- (51) International Patent Classification⁷: **C12N 15/09**, (74) Common Representative: **GENSET**; Intellectual Property Department, 24, rue Royale, F-75008 Paris (FR).
C07K 14/47
- (21) International Application Number: PCT/IB00/01938 (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (22) International Filing Date: 7 December 2000 (07.12.2000)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/169,629 8 December 1999 (08.12.1999) US
60/187,470 6 March 2000 (06.03.2000) US
- (71) Applicant (*for all designated States except US*): **GENSET** [FR/FR]; Intellectual Property Department, 24, rue Royale, F-75008 Paris (FR).
- (72) Inventors; and
(75) Inventors/Applicants (*for US only*): **DUMAS MILNE EDWARDS, Jean-Baptiste** [FR/FR]; 8, rue Grégoire de Tours, F-75006 Paris (FR). **BOUGUELERET, Lydie** [FR/FR]; 108, avenue Victor Hugo, F-92170 Vanves (FR). **JOBERT, Séverin** [FR/FR]; 7, impasse Tourneux, F-75010 Paris (FR).
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— Without international search report and to be republished upon receipt of that report.

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: FULL-LENGTH HUMAN cDNAs ENCODING POTENTIALLY SECRETED PROTEINS

(57) Abstract: The invention concerns GENSET polynucleotides and polypeptides. Such GENSET products may be used as reagents in forensic analyses, as chromosome markers, as tissue/cell/organelle-specific markers, in the production of expression vectors. In addition, they may be used in screening and diagnosis assays for abnormal GENSET expression and/or biological activity and for screening compounds that may be used in the treatment of GENSET-related disorders.

WO 01/42451 A2

Full-length human cDNAs encoding potentially secreted proteins

Related application

The present application claims priority to the US Provisional Patent Applications Serial Nos 60/169,629 and 60/187,470 filed December, 8, 1999, and March, 6, 2000, respectively, the
5 disclosures of which are incorporated herein by reference in their entireties.

Field of the invention

The present invention is directed to polynucleotides encoding GENSET polypeptides, fragments thereof, and the regulatory regions located in the 5'- and 3'-ends of the GENSET genes. The invention also concerns polypeptides encoded by the GENSET polynucleotides and fragments
10 thereof. The present invention also relates to recombinant vectors, which include the polynucleotides of the present invention, particularly recombinant vectors comprising a GENSET regulatory region or a sequence encoding a GENSET polypeptide, and to host cells containing the polynucleotides of the invention, as well as to methods of making such vectors and host cells. The present invention further relates to the use of these recombinant vectors and host cells in the
15 production of the polypeptides of the invention. The invention further relates to antibodies that specifically bind to the polypeptides of the invention and to methods for producing such antibodies and fragments thereof. The invention also provides for methods of detecting the presence of the polynucleotides and polypeptides of the present invention in a sample, methods of diagnosis and screening of abnormal GENSET gene expression and/or biological activity, methods of screening
20 compounds for their ability to modulate the activity or expression of GENSET genes and uses of such compounds.

Background of the invention

The estimated 50,000-100,000 genes scattered along the human chromosomes offer tremendous promise for the understanding, diagnosis, and treatment of human diseases. In addition,
25 probes capable of specifically hybridizing to loci distributed throughout the human genome find applications in the construction of high resolution chromosome maps and in the identification of individuals.

Currently, two different approaches are being pursued for identifying and characterizing the genes distributed along the human genome. In one approach, large fragments of genomic DNA are
30 isolated, cloned, and sequenced. Potential open reading frames in these genomic sequences are identified using bio-informatics software. However, this approach entails sequencing large stretches of human DNA which do not encode proteins in order to find the protein encoding sequences scattered throughout the genome. In addition to requiring extensive sequencing, the bio-

informatics software may mischaracterize the genomic sequences obtained, *i.e.*, labeling non-coding DNA as coding DNA and vice versa.

An alternative approach takes a more direct route to identifying and characterizing human genes. In this approach, complementary DNAs (cDNAs) are synthesized from isolated messenger RNAs (mRNAs) which encode human proteins. Using this approach, sequencing is only performed on DNA which is derived from protein coding fragments of the genome. In the past, these cDNAs, after short EST sequences were obtained from oligo-dT primed cDNA libraries. Accordingly, they mainly corresponded to the 3' untranslated region of the mRNA. In part, the prevalence of EST sequences derived from the 3' end of the mRNA is a result of the fact that typical techniques for obtaining cDNAs, are not well suited for isolating cDNA sequences derived from the 5' ends of mRNAs (Adams *et al.*, *Nature* 377:3-174, 1996, Hillier *et al.*, *Genome Res.* 6:807-828, 1996). In addition, in those reported instances where longer cDNA sequences have been obtained, the reported sequences typically correspond to coding sequences and do not include the full 5' untranslated region (5'UTR) of the mRNA from which the cDNA is derived. Indeed, 5'UTRs have been shown to affect either the stability or translation of mRNAs. Thus, regulation of gene expression may be achieved through the use of alternative 5'UTRs as shown, for instance, for the translation of the tissue inhibitor of metalloprotease mRNA in mitogenically activated cells (Waterhouse *et al.*, *J Biol Chem.* 265:5585-9, 1990). Furthermore, modification of 5'UTR through mutation, insertion or translocation events may even be implied in pathogenesis. For instance, the fragile X syndrome, the most common cause of inherited mental retardation, is partly due to an insertion of multiple CGG trinucleotides in the 5'UTR of the fragile X mRNA resulting in the inhibition of protein synthesis via ribosome stalling (Feng *et al.*, *Science* 268:731-4, 1995). An aberrant mutation in regions of the 5'UTR known to inhibit translation of the proto-oncogene *c-myc* was shown to result in upregulation of c-myc protein levels in cells derived from patients with multiple myelomas (Willis *et al.*, *Curr Top Microbiol Immunol* 224:269-76, 1997). In addition, the use of oligo-dT primed cDNA libraries does not allow the isolation of complete 5'UTRs since such incomplete sequences obtained by this process may not include the first exon of the mRNA, particularly in situations where the first exon is short. Furthermore, they may not include some exons, often short ones, which are located upstream of splicing sites. Thus, there is a need to obtain sequences derived from the 5' ends of mRNAs.

Moreover, despite the great amount of EST data that large-scale sequencing projects have yielded (Adams *et al.*, *Nature* 377:174, 1996, Hillier *et al.*, *Genome Res.* 6:807-828, 1996), information concerning the biological function of the mRNAs corresponding to such obtained cDNAs has revealed to be limited. Indeed, whereas the knowledge of the complete coding sequence is absolutely necessary to investigate the biological function of mRNAs, ESTs yield only partial coding sequences. So far, large-scale full-length cDNA cloning has been achieved only with limited success because of the poor efficiency of methods for constructing full-length cDNA

libraries. Indeed, such methods require either a large amount of mRNA (Ederly *et al.*, 1995), thus resulting in non representative full-length libraries when small amounts of tissue are available or require PCR amplification (Maruyama *et al.*, 1994; CLONTECHniques, 1996) to obtain a reasonable number of clones, thus yielding strongly biased cDNA libraries where rare and long
5 cDNAs are lost. Thus, there is a need to obtain full-length cDNAs, *i.e.* cDNAs containing the full coding sequence of their corresponding mRNAs. The present application presents a number of cDNAs, called GENSET polynucleotides, isolated from full-length cDNA libraries obtained from the methods described in PCT publication WO 00/37491.

While many sequences derived from human chromosomes have practical applications,
10 approaches based on the identification and characterization of those chromosomal sequences which encode a protein product are particularly relevant to diagnostic and therapeutic uses. Of the 50,000-100,000 protein coding genes, those genes encoding proteins which are secreted from the cell in which they are synthesized, as well as the secreted proteins themselves, are particularly valuable as potential therapeutic agents. Such proteins are often involved in cell to cell communication and
15 may be responsible for producing a clinically relevant response in their target cells. In fact, several secretory proteins, including tissue plasminogen activator, G-CSF, GM-CSF, erythropoietin, human growth hormone, insulin, interferon- α , interferon- β , interferon- γ , and interleukin-2, are currently in clinical use. These proteins are used to treat a wide range of conditions, including acute myocardial infarction, acute ischemic stroke, anemia, diabetes, growth hormone deficiency, hepatitis, kidney
20 carcinoma, chemotherapy induced neutropenia and multiple sclerosis. For these reasons, cDNAs encoding secreted proteins or fragments thereof represent a particularly valuable source of therapeutic agents. Thus, there is a need for the identification and characterization of secreted proteins and the nucleic acids encoding them.

In addition to being therapeutically useful themselves, secretory proteins include short
25 peptides, called signal peptides, at their amino termini which direct their secretion. These signal peptides are encoded by the signal sequences located at the 5' ends of the coding sequences of genes encoding secreted proteins. Because these signal peptides will direct the extracellular secretion of any protein to which they are operably linked, the signal sequences may be exploited to direct the efficient secretion of any protein by operably linking the signal sequences to a gene encoding the
30 protein for which secretion is desired. In addition, fragments of the signal peptides called membrane-translocating sequences, may also be used to direct the intracellular import of a peptide or protein of interest. This may prove beneficial in gene therapy strategies in which it is desired to deliver a particular gene product to cells other than the cells in which it is produced. Signal sequences encoding signal peptides also find application in simplifying protein purification
35 techniques. In such applications, the extracellular secretion of the desired protein greatly facilitates purification by reducing the number of undesired proteins from which the desired protein must be

selected. Thus, there exists a need to identify and characterize the 5' fragments of the genes for secretory proteins which encode signal peptides.

Sequences coding for human proteins may also find application as therapeutics or diagnostics. In particular, such sequences may be used to determine whether an individual is likely to express a detectable phenotype, such as a disease, as a consequence of a mutation in the coding sequence for a protein. In instances where the individual is at risk of suffering from a disease or other undesirable phenotype as a result of a mutation in such a coding sequence, the undesirable phenotype may be corrected by introducing a normal coding sequence using gene therapy. Alternatively, if the undesirable phenotype results from overexpression of the protein encoded by the coding sequence, expression of the protein may be reduced using antisense or triple helix based strategies.

The GENSET human polypeptides encoded by the coding sequences may also be used as therapeutics by administering them directly to an individual having a condition, such as a disease, resulting from a mutation in the sequence encoding the polypeptide. In such an instance, the condition can be cured or ameliorated by administering the polypeptide to the individual.

In addition, the human polypeptides or fragments thereof may be used to generate antibodies useful in determining the tissue type or species of origin of a biological sample. The antibodies may also be used to determine the subcellular localization of the human polypeptides or the cellular localization of polypeptides which have been fused to the human polypeptides. In addition, the antibodies may also be used in immunoaffinity chromatography techniques to isolate, purify, or enrich the human polypeptide or a target polypeptide which has been fused to the human polypeptide.

Public information on the number of human genes for which the promoters and upstream regulatory regions have been identified and characterized is quite limited. In part, this may be due to the difficulty of isolating such regulatory sequences. Upstream regulatory sequences such as transcription factor binding sites are typically too short to be utilized as probes for isolating promoters from human genomic libraries. Recently, some approaches have been developed to isolate human promoters. One of them consists of making a CpG island library (Cross *et al.*, *Nature Genetics* 6: 236-244, 1994). The second consists of isolating human genomic DNA sequences containing SpeI binding sites by the use of SpeI binding protein. (Mortlock *et al.*, *Genome Res.* 6:327-335, 1996). Both of these approaches have their limits due to a lack of specificity and of comprehensiveness. Thus, there exists a need to identify and systematically characterize the 5' fragments of the genes.

cDNAs including the 5' ends of their corresponding mRNA may be used to efficiently identify and isolate 5'UTRs and upstream regulatory regions which control the location, developmental stage, rate, and quantity of protein synthesis, as well as the stability of the mRNA (Theil *et al.*, *BioFactors* 4:87-93, (1993). Once identified and characterized, these regulatory

regions may be utilized in gene therapy or protein purification schemes to obtain the desired amount and locations of protein synthesis or to inhibit, reduce, or prevent the synthesis of undesirable gene products.

In addition, cDNAs containing the 5' ends of protein genes may include sequences useful as probes for chromosome mapping and the identification of individuals. Thus, there is a need to identify and characterize the sequences upstream of the 5' coding sequences of genes encoding proteins.

Summary of the invention

The present invention provides compositions containing a purified or isolated polynucleotide comprising, consisting of, or consisting essentially of a nucleotide sequence selected from the group consisting of: (a) the sequences of SEQ ID Nos: 1-241; (b) the sequences of clone inserts of the deposited clone pool; (c) the full coding sequences of SEQ ID Nos: 1-241; (d) the full coding sequences of the clone inserts of the deposited clone pool; (e) the sequences encoding one of the polypeptides of SEQ ID Nos: 242-482; (f) the sequences encoding one of the polypeptides encoded by the clone inserts of the deposited clone pool; (g) the genomic sequences coding for GENSET polypeptides; (h) the 5' transcriptional regulatory regions of GENSET genes; (i) the 3' transcriptional regulatory regions of GENSET genes; (j) the polynucleotides comprising the nucleotide sequence of any combination of (g)-(i); (k) the variant polynucleotides of any of the polynucleotides of (a)-(j); (l) the polynucleotides comprising a nucleotide sequence of (a)-(k), wherein the polynucleotide is single stranded, double stranded, or a portion is single stranded and a portion is double stranded; (m) the polynucleotides comprising a nucleotide sequence complementary to any of the single stranded polynucleotides of (l). The invention further provides for fragments of the nucleic acid molecules of (a)-(m) described above.

The present invention also provides biologically active forms, variants, fragments and derivatives of the present proteins, where "biologically active" indicates that the form, variant, fragment, or derivative, has any detectable activity in any in vitro assay known in the art or described herein, or has any detectable function in vivo. In preferred embodiments, a determination of whether a particular polypeptide is biologically active will be made based on any of the specific assays or functional characteristics provided below for each of the proteins of this invention.

Therefore, one embodiment of the present invention is a composition containing a purified or isolated nucleic acid comprising a sequence selected from the group consisting of sequences of SEQ ID NOS: 1-241 and sequences of clone inserts of the deposited clone pool, sequences complementary thereto, allelic variants thereof, and degenerate variants thereof. In one aspect of this embodiment, the nucleic acid is recombinant.

Another embodiment of the present invention is a composition containing a purified or isolated nucleic acid comprising at least 8 consecutive nucleotides of a sequence selected from the

group consisting of sequences of SEQ ID NOs: 1-241 and sequences of clone inserts of the deposited clone pool, sequences complementary thereto, allelic variants thereof, and degenerate variants thereof. In one aspect of this embodiment, the nucleic acid comprises at least 10, 12, 15, 18, 20, 25, 28, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, 500, 800, 1000, 1500, or 2000

5 consecutive nucleotides of said selected sequence, sequences complementary thereto, allelic variants thereof, and degenerate variants thereof. The nucleic acid may be a recombinant nucleic acid.

Another embodiment of the present invention is a composition comprising a vertebrate purified or isolated nucleic acid of at least 15, 18, 20, 23, 25, 28, 30, 35, 40, 50, 75, 100, 200, 300, 500, 1000 or 2000 nucleotides in length which hybridizes under stringent conditions to any
10 polynucleotide of the invention, preferably a sequence selected from the group consisting of sequences of SEQ ID NOs: 1-241 and sequences of clone inserts of the deposited clone pool, sequences complementary thereto. In one aspect of this embodiment, the nucleic acid is recombinant.

Another embodiment of the present invention is a composition containing a purified or
15 isolated nucleic acid comprising the full coding sequences of a sequence selected from the group consisting of sequences of SEQ ID NOs: 1-241 and sequences of clone inserts of the deposited clone pool, or an allelic variant thereof. In one aspect of this embodiment, the nucleic acid is recombinant.

A further embodiment of the present invention is a composition containing a purified or
20 isolated nucleic acid comprising a contiguous span of a sequence selected from the group consisting of sequences of SEQ ID NOs: 1-31 and 33-143 and sequences of clone inserts encoding secreted proteins in the deposited clone pool, or an allelic variant thereof, wherein said contiguous span encodes a mature protein. In one aspect of this embodiment, the nucleic acid is recombinant. In another aspect of this embodiment, the nucleic acid is an expression vector wherein said contiguous
25 span which encodes a mature protein is operably linked to a promoter.

Yet another embodiment of the present invention is a composition containing a purified or isolated nucleic acid comprising a contiguous span of a sequence selected from the group consisting of sequences of SEQ ID NOs: 1-31 and 33-143 and sequences of clone inserts encoding secreted proteins in the deposited clone pool, or an allelic variant thereof, wherein said contiguous span
30 encodes a signal peptide. In one aspect of this embodiment, the nucleic acid is recombinant. In another aspect of this embodiment, the nucleic acid is an fusion vector wherein said contiguous span which encodes a signal peptide is operably linked to a second nucleic acid encoding an heterologous polypeptide.

Another embodiment of the present invention is a composition containing a purified or
35 isolated nucleic acid encoding a polypeptide comprising a sequence selected from the group consisting of sequences of SEQ ID NOs: 1-241 and sequences of clone inserts of the deposited

clone pool, or allelic variant thereof. In one aspect of this embodiment, the nucleic acid is recombinant.

Another embodiment of the present invention is a composition containing a purified or isolated nucleic acid encoding a polypeptide comprising the sequence of a mature protein included
5 in a sequence selected from the group consisting of sequences of SEQ ID NOs: 1-31 and 33-143 and sequences of clone inserts encoding secreted proteins in the deposited clone pool, or allelic variant thereof. In one aspect of this embodiment, the nucleic acid is recombinant.

Another embodiment of the present invention is a composition containing a purified or isolated nucleic acid encoding a polypeptide comprising the sequence of a signal peptide included
10 in a sequence selected from the group consisting of sequences of SEQ ID NOs: 1-31 and 33-143 and sequences of clone inserts encoding secreted proteins in the deposited clone pool, or allelic variant thereof. In another aspect it is present in a vector of the invention.

Further embodiments of the invention include compositions containing purified or isolated polynucleotides that comprise, a nucleotide sequence at least 70% identical, more preferably at least
15 75% identical, and still more preferably at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identical to any of the polynucleotides of the present invention. Methods of determining identity include those well known in the art and described herein. Such analyses can be performed using a full length polynucleotide sequence or using a subsequence of any length. For example, any two sequences can be compared over a region, in either protein or in both proteins, of any 10, 25, 50,
20 100, 250, 500, 1000, 2000 or more contiguous nucleotides. In addition, any two sequences can be identified as homologous even when they share sequence homology over a limited region of either polynucleotide, for example over a region of at least about 10, 25, 50, 100, 250, 500, 1000, or more contiguous nucleotides.

The invention further provides compositions containing a purified or isolated polypeptide
25 comprising, consisting of, or consisting essentially of an amino acid sequence selected from the group consisting of: (a) the polypeptides of SEQ ID Nos: 242-482; (b) the polypeptides encoded by the clone inserts of the deposited clone pool; (c) the epitope-bearing fragments of the polypeptides of SEQ ID Nos: 242-482; (d) the epitope-bearing fragments of the polypeptides encoded by the clone inserts contained in the deposited clone pool; (e) the domains of the polypeptides of SEQ ID
30 Nos: 242-482; (f) the domains of the polypeptides encoded by the clone inserts contained in the deposited clone pool; and (g) the allelic variant polypeptides of any of the polypeptides of (a)-(f). The invention further provides for fragments of the polypeptides of (a)-(g) above, such as those having biological activity or comprising biologically functional domain(s).

Yet another embodiment of the present invention is a composition containing a purified or
35 isolated protein comprising a sequence selected from the group consisting of sequences of SEQ ID NOs: 242-482 and sequences of polypeptides encoded by clone inserts of the deposited clone pool, or allelic variant thereof.

Another embodiment of the present invention is a composition containing a purified or isolated polypeptide comprising at least 5, 6 or 8 consecutive amino acids of a sequence selected from the group consisting of sequences of SEQ ID NOs: 242-482 and sequences of polypeptides encoded by clone inserts of the deposited clone pool, or allelic variant thereof. In one aspect of this
5 embodiment, the purified or isolated polypeptide comprises at least 10, 12, 15, 20, 25, 30, 35, 40, 50, 60, 75, 100, 150, 200, 250, 300, 350, 400, 450 or 500 consecutive amino acids of said selected sequence or allelic variant thereof.

Another embodiment of the present invention is a composition containing an isolated or purified polypeptide comprising a signal peptide of a sequence selected from the group consisting
10 of sequences of SEQ ID NOs: 242-272 and 274-384 and sequences of polypeptides encoded by clone inserts of the deposited clone pool, or allelic variant thereof.

Yet another embodiment of the present invention is a composition containing an isolated or purified polypeptide comprising a mature protein of a sequence selected from the group consisting of sequences of SEQ ID NOs: 242-272 and 274-384 and sequences of polypeptides encoded by
15 clone inserts of the deposited clone pool, or allelic variant thereof.

A further embodiment of the present invention are compositions containing polypeptide having an amino acid sequence with at least 70% similarity, and more preferably at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% similarity to a polypeptide of the present invention, as well as polypeptides having an amino acid sequence at least 70% identical, more preferably at least
20 75% identical, and still more preferably 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identical to a polypeptide of the present invention. Such analyses can be performed using a full length polypeptide sequence or using a subsequence of any length. For example, any two sequences can be compared over a region, in either protein or in both proteins, of any 10, 25, 50, 100, 250, 500, 1000, 2000 or more contiguous amino acids. In addition, any two sequences can be identified as
25 homologous even when they share sequence homology over a limited region of either protein, for example over a region of at least about 10, 25, 50, 100, 250, 500, 1000, or more contiguous amino acids. Further included in the invention are compositions comprising a purified or isolated nucleic acid molecule encoding such polypeptides. Methods for determining identity include those well known in the art and described herein.

30 The present invention also relates to compositions comprising recombinant vectors, which include the purified or isolated polynucleotides of the present invention, and to host cells recombinant for the polynucleotides of the present invention, as well as to methods of making such vectors and host cells. The present invention further relates to the use of these recombinant vectors and recombinant host cells in the production of GENSET polypeptides.

35 Consequently, another embodiment of the invention is a vector comprising any polynucleotide of the invention. In a preferred embodiment, the vector is an expression vector comprising a nucleic acid sequence encoding a polypeptide selected from the group consisting of

sequences of SEQ ID NOs: 242-482 and sequences of polypeptides encoded by the clone inserts of the deposited clone pool, or allelic variant thereof, wherein said nucleic acid sequence is operably linked to a promoter. In another preferred embodiment, the vector is a secretion vector comprising a nucleic acid sequence encoding a signal peptide selected from the group consisting of signal peptides of sequences of SEQ ID NOs: 242-272 and 274-384 and sequences of secreted polypeptides encoded by the clone inserts of the deposited clone pool, or allelic variant thereof, wherein said nucleic acid sequence is operably linked to an heterologous protein such that said signal peptide will direct the secretion of said heterologous protein.

A further embodiment of the present invention is a method of making a protein comprising a sequence selected from the group consisting of sequences of SEQ ID NOs: 242-482 and sequences of polypeptides encoded by clone inserts of the deposited clone pool, comprising the steps of

- a) obtaining a cDNA comprising a sequence selected from the group consisting of sequences of SEQ ID NOs: 1-241 and sequences of clone inserts of the deposited clone pool;
- b) inserting said cDNA in an expression vector such that said cDNA is operably linked to a promoter; and
- c) introducing said expression vector into a host cell whereby the host cell produces the protein encoded by said cDNA.

In one aspect of this embodiment, the method further comprises the step of isolating said protein.

Another embodiment of the present invention is a protein obtainable by the method described in the preceding paragraph.

Another embodiment of the present invention is a method of making a protein comprising the amino acid sequence of the mature protein contained in a sequence selected from the group consisting of sequences of SEQ ID NOs: 242-272 and 274-384 and sequences of polypeptides encoded by clone inserts of the deposited clone pool, comprising the steps of

- a) obtaining a cDNA comprising a sequence selected from the group consisting of sequences of SEQ ID NOs: 1-31 and 33-143 and sequences of clone inserts of the deposited clone pool, wherein said cDNA encodes a mature protein;
- b) inserting said cDNA in an expression vector such that said cDNA is operably linked to a promoter; and
- c) introducing said expression vector into a host cell whereby the host cell produces the mature protein encoded by said cDNA.

In one aspect of this embodiment, the method further comprises the step of isolating said protein.

Another embodiment of the present invention is a mature protein obtainable by the method described in the preceding paragraph.

Another embodiment of the present invention is a composition containing a host cell containing the purified or isolated nucleic acids comprising a sequence selected from the group consisting of sequences of SEQ ID NOs: 1-241 and sequences of clone inserts of the deposited clone pool or a sequence complementary thereto described herein.

- 5 Another embodiment of the present invention is a composition containing a host cell containing the purified or isolated nucleic acids comprising the full coding sequences of a sequence selected from the group consisting of sequences of SEQ ID NOs: 1-241 and sequences of clone inserts of the deposited clone pool.

- Another embodiment of the present invention is a composition containing a host cell
10 containing the purified or isolated nucleic acids comprising a contiguous span of a sequence selected from the group consisting of sequences of SEQ ID NOs: 1-31 and 33-143 and sequences of clone inserts of the deposited clone pool, wherein said contiguous span codes for a mature protein.

- Another embodiment of the present invention is a composition containing a host cell containing the purified or isolated nucleic acids comprising a contiguous span of a sequence
15 selected from the group consisting of sequences of SEQ ID NOs: 1-31 and 33-143 and sequences of clone inserts of the deposited clone pool, wherein said contiguous span codes for a signal peptide.

The invention further relates to other methods of making the polypeptides of the present invention.

- The present invention further relates to transgenic plants or animals, wherein said transgenic
20 plant or animal is transgenic for a polynucleotide of the present invention and expresses a polypeptide of the present invention.

The invention further relates to compositions comprising antibodies that specifically bind to the GENSET polypeptides of the present invention and fragments thereof as well as to methods for producing such antibodies and fragments thereof.

- 25 Therefore, another embodiment of the present invention is a composition containing a purified or isolated antibody capable of specifically binding to a protein comprising a sequence selected from the group consisting of sequences of SEQ ID NOs: 242-482 and sequences of polypeptides encoded by clone inserts of the deposited clone pool. In one aspect of this embodiment, the antibody is capable of binding to a polypeptide comprising at least 6 consecutive
30 amino acids, at least 8 consecutive amino acids, or at least 10 consecutive amino acids of said selected sequence.

- The invention also provides kits and methods of detecting GENSET gene expression and/or biological activity in a biological sample. One such method involves assaying for the expression of a GENSET polynucleotide in a biological sample using polymerase chain reaction (PCR) to amplify
35 and detect GENSET polynucleotides or Southern and Northern blot hybridization to detect GENSET genomic DNA, cDNA or mRNA. Alternatively, a method of detecting GENSET gene

expression in a test sample can be accomplished using a compound which binds to a GENSET polypeptide of the present invention or a portion of a GENSET polypeptide.

The present invention also relates to diagnostic methods of identifying individuals or non-human animals having elevated or reduced levels of GENSET products, which individuals are likely to benefit from therapies to suppress or enhance GENSET gene expression, respectively and to methods of identifying individuals or non-human animals at increased risk for developing, or present state of having, certain diseases/disorders associated with GENSET gene abnormal expression or biological activity.

The present invention also relates to kits and methods of screening compounds for their ability to modulate (e.g. increase or inhibit) the activity or expression of GENSET genes including compounds that interact with GENSET gene regulatory sequences and compounds that interact directly or indirectly with GENSET polypeptides. Uses of such compounds are also under the scope of the present invention.

The present invention also relates to pharmaceutical or physiologically acceptable compositions comprising, an active agent, the polypeptides, polynucleotides or antibodies of the present invention.

The present invention also relates to computer systems containing cDNA codes and polypeptides codes of sequences of the invention and to computer-related methods of comparing sequences, identifying homology or features using GENSET sequences of the invention.

In another aspect, the present invention provides an isolated polynucleotide, said polynucleotide comprising a nucleic acid sequence encoding i) a polypeptide comprising an amino acid sequence having at least about 80% identity to any one of the sequences shown as SEQ ID NOs:242-482 or any one of the sequences of polypeptides encoded by the clone inserts of the deposited clone pool; or a biologically active fragment of said polypeptide.

In one embodiment, the polypeptide comprises any one of the sequences shown as SEQ ID NOs:242-482 or any one of the sequences of the polypeptides encoded by the clone inserts of the deposited clone pool. In another embodiment, the polypeptide comprises a signal peptide. In another embodiment, the polypeptide is a mature protein. In another embodiment, the nucleic acid sequence has at least about 80% identity over at least about 100 contiguous nucleotides to any one of the sequences shown as SEQ ID NOs:1-241 or any one of the sequences of the clone inserts of the deposited clone pool. In another embodiment, the polynucleotide hybridizes under stringent conditions to a polynucleotide comprising any one of the sequences shown as SEQ ID NOs:1-241 or any one of the sequences of the clone inserts of the deposited clone pool. In another embodiment, the nucleic acid sequence comprises any one of the sequences shown as SEQ ID NOs:1-241 or any one the sequences of the clone inserts of the deposited clone pool. In another embodiment, the polynucleotide is operably linked to a promoter.

In another aspect, the present invention provides an expression vector comprising the polynucleotide operably linked to a promoter. In another aspect, the present invention provides a host cell recombinant for the polynucleotide. In another aspect, the present invention provides a non-human transgenic animal comprising the host cell.

5 In another aspect, the present invention provides a method of making a GENSET polypeptide, the method comprising a) providing a population of host cells comprising a herein-described polynucleotide and b) culturing the population of host cells under conditions conducive to the production of the polypeptide within said host cells.

In one embodiment, the method further comprises purifying the polypeptide from the
10 population of host cells.

In another aspect, the present invention provides a method of making a GENSET polypeptide, the method comprising a) providing a population of cells comprising a herein-described polynucleotide; b) culturing the population of cells under conditions conducive to the production of the polypeptide within the cells; and c) purifying the polypeptide from the population
15 of cells.

In another aspect, the present invention provides an isolated polynucleotide, the polynucleotide comprising a nucleic acid sequence having at least about 80% identity over at least about 100 contiguous nucleotides to any one of the sequences shown as SEQ ID NOs:1-241 or any one of the sequences of the clone inserts of the deposited clone pool.

20 In one embodiment, the polynucleotide hybridizes under stringent conditions to a polynucleotide comprising any one of the sequences shown as SEQ ID NOs:1-241 or any one of the sequences of the clone inserts of the deposited clone pool. In another embodiment, the polynucleotide comprises any one of the sequences shown as SEQ ID NOs:1-241 or any one of the sequences of the clone inserts of the deposited clone pool.

25 In another aspect, the present invention provides a biologically active polypeptide encoded by any of the herein-described polynucleotides.

In another aspect, the present invention provides an isolated polypeptide or biologically active fragment thereof, the polypeptide comprising an amino acid sequence having at least about 80% sequence identity to any one of the sequences shown as SEQ ID NOs:242-482 or any one of
30 the sequences of polypeptides encoded by the clone inserts of the deposited clone pool.

In one embodiment, the polypeptide is selectively recognized by an antibody raised against an antigenic polypeptide, or an antigenic fragment thereof, said antigenic polypeptide comprising any one of the sequences shown as SEQ ID NOs:242-482 or any one of the sequences of polypeptides encoded by the clone inserts of the deposited clone pool. In another embodiment, the
35 polypeptide comprises any one of the sequences shown as SEQ ID NOs:242-482 or any one of the sequences of polypeptides encoded by the clone inserts of the deposited clone pool. In another

embodiment, the polypeptide comprises a signal peptide. In another embodiment, the polypeptide is a mature protein.

In another aspect, the present invention provides an antibody that specifically binds to any of the herein-described polypeptides.

5 In another aspect, the present invention provides a method of determining whether a GENSET gene is expressed within a mammal, the method comprising the steps of: a) providing a biological sample from said mammal; b) contacting said biological sample with either of: i) a polynucleotide that hybridizes under stringent conditions to the polynucleotide of claim 1; or ii) a polypeptide that specifically binds to the polypeptide of claim 19; and c) detecting the presence or
10 absence of hybridization between the polynucleotide and an RNA species within the sample, or the presence or absence of binding of the polypeptide to a protein within the sample; wherein a detection of the hybridization or of the binding indicates that the GENSET gene is expressed within the mammal.

In one embodiment, the polynucleotide is a primer, and the hybridization is detected by
15 detecting the presence of an amplification product comprising the sequence of the primer. In another embodiment, the polypeptide is an antibody.

In another aspect, the present invention provides a method of determining whether a mammal has an elevated or reduced level of GENSET gene expression, the method comprising the steps of : a) providing a biological sample from the mammal; and b) comparing the amount of any
20 of the herein-described polypeptides, or of an RNA species encoding the polypeptide, within the biological sample with a level detected in or expected from a control sample; wherein an increased amount of the polypeptide or the RNA species within the biological sample compared to the level detected in or expected from the control sample indicates that the mammal has an elevated level of the GENSET gene expression, and wherein a decreased amount of the polypeptide or the RNA
25 species within the biological sample compared to the level detected in or expected from the control sample indicates that the mammal has a reduced level of the GENSET gene expression.

In another aspect, the present invention provides a method of identifying a candidate modulator of a GENSET polypeptide, the method comprising : a) contacting any of the herein-described polypeptides with a test compound; and b) determining whether the compound
30 specifically binds to the polypeptide; wherein a detection that the compound specifically binds to the polypeptide indicates that the compound is a candidate modulator of the GENSET polypeptide.

Brief description of drawings

Figure 1 is a map of the expression vector pPT

35 Figure 2 is a block diagram of an exemplary computer system.

Figure 3 is a flow diagram illustrating one embodiment of a process 200 for comparing a new nucleotide or protein sequence with a database of sequences in order to determine the identity levels between the new sequence and the sequences in the database.

Figure 4 is a flow diagram illustrating one embodiment of a process 250 in a computer for determining whether two sequences are homologous.

Figure 5 is a flow diagram illustrating one embodiment of an identifier process 300 for detecting the presence of a feature in a sequence.

Brief Description of Tables

Table I provides the applicant's internal designation number assigned to each sequence identification number and indicates whether the sequence is a nucleic acid sequence or a polypeptide sequence, and in which vector the cDNA was cloned.

Table II provides structural features for each cDNA of SEQ ID Nos: 1-241 i.e., the locations of the full coding sequences, the signal peptides, the mature polypeptides, the polyA signal and the polyA site.

Table III lists variants for cDNAs of the present invention.

Table IV provides the positions of fragments which are preferably excluded from the present invention.

Tables Va and b provides the positions of fragments which are preferably excluded or included in the present invention. Table IV and Tables Va, and Table Vb provide for the inclusion and exclusion of polynucleotides independently from each other in addition to those described elsewhere in the specification and is therefore, not meant as limiting description.

Table VI lists known biologically structural and functional domains for the polypeptides of the present invention.

Table VII lists antigenic peaks of predicted antigenic epitopes for polypeptides of the present invention.

Table VIII lists the putative chromosomal location of the polynucleotides of the present invention.

Table IX list the Genset's cDNA libraries of tissues and cell types examined that express the polynucleotides of the present invention.

Table X relates to the bias in spatial distribution of the polynucleotide sequences of the present invention.

Table XI lists predicted subcellular localization for cDNAs of the present invention.

Table XII gives the correspondence between the polynucleotides of the US priority applications, namely the US Provisional Patent Applications Serial Nos 60/169,629 and 60/187, (column entitled "Seq Id No in priority applications") and the polynucleotides of the present application (column entitled "Seq Id No in present application").

Brief description of sequence listing

SEQ ID Nos: 1-31 and 33-143 are the nucleotide sequences of cDNAs encoding a potentially secreted protein. The locations of the ORFs and sequences encoding signal peptides are listed in the accompanying Sequence Listing. In addition, the von Heijne score of the signal peptide computed as described below is listed as the "score" in the accompanying Sequence Listing. The sequence of the signal-peptide is listed as "seq" in the accompanying Sequence Listing. The "/" in the signal peptide sequence indicates the location where proteolytic cleavage of the signal peptide occurs to generate a mature protein. When appropriate, the locations of the first and last nucleotides of the coding sequences, eventually the locations of the first and last nucleotides of the polyA and the locations of the first and last nucleotides of the polyA sites are indicated.

SEQ ID Nos. 32 and 144-241 are the nucleotide sequences of cDNAs in which no sequence encoding a signal peptide has been identified to date. However, it remains possible that subsequent analysis will identify a sequence encoding a signal peptide in these nucleic acids. The locations of the ORFs are listed in the accompanying Sequence Listing. When appropriate, the locations of the first and last nucleotides of the coding sequences, eventually the locations of the first and last nucleotides of the polyA and the locations of the first and last nucleotides of the polyA sites are indicated.

SEQ ID Nos: 242-272 and 274-384 are the amino acid sequences of polypeptides which contain a signal peptide. These polypeptides are encoded by the cDNAs of SEQ ID Nos: 1-31 and 33-143 respectively. The location of the signal peptide is listed in the accompanying Sequence Listing.

SEQ ID Nos: 273 and 385-482 are the amino acid sequences of polypeptides in which no signal peptide has been identified to date. However, it remains possible that subsequent analysis will identify a signal peptide in these polypeptides. These polypeptides are encoded by the nucleic acids of SEQ ID Nos: 32 and 144-241 respectively.

In accordance with the regulations relating to Sequence Listings, the following codes have been used in the Sequence Listing to describes nucleotide sequences. The code "r" in the sequences indicates that the nucleotide may be a guanine or an adenine. The code "y" in the sequences indicates that the nucleotide may be a thymine or a cytosine. The code "m" in the sequences indicates that the nucleotide may be an adenine or a cytosine. The code "k" in the sequences indicates that the nucleotide may be a guanine or a thymine. The code "s" in the sequences indicates that the nucleotide may be a guanine or a cytosine. The code "w" in the sequences indicates that the nucleotide may be an adenine or an thymine. In addition, all instances of the symbol "n" in the nucleic acid sequences mean that the nucleotide can be adenine, guanine, cytosine or thymine.

In some instances, the polypeptide sequences in the Sequence Listing contain the symbol "Xaa." These "Xaa" symbols indicate either (1) a residue which cannot be identified because of nucleotide sequence ambiguity or (2) a stop codon in the determined sequence where applicants

believe one should not exist (if the sequence were determined more accurately). In some instances, several possible identities of the unknown amino acids may be suggested by the genetic code.

In the case of secreted proteins, it should be noted that, in accordance with the regulations governing Sequence Listings, in the appended Sequence Listing, the encoded protein (i.e. the protein containing the signal peptide and the mature protein or part thereof) extends from an amino acid residue having a negative number through a positively numbered amino acid residue. Thus, the first amino acid of the mature protein resulting from cleavage of the signal peptide is designated as amino acid number 1, and the first amino acid of the signal peptide is designated with the appropriate negative number. However, in the present application, positions on amino acid sequences are always given on the full length polypeptide, the first amino acid of the signal peptide being designated as amino acid number 1.

Detailed description

DEFINITIONS

Before describing the invention in greater detail, the following definitions are set forth to illustrate and define the meaning and scope of the terms used to describe the invention herein.

The terms "GENSET gene", when used herein, encompasses genomic, mRNA and cDNA sequences encoding the GENSET protein, including the 5' and 3' untranslated regions of said sequences.

As used herein, a "secreted" protein is one which, when expressed in a suitable host cell, is transported across or through a membrane, including transport as a result of signal peptides in its amino acid sequence. "Secreted" proteins include without limitation proteins secreted wholly (e.g. soluble proteins), or partially (e.g. receptors) from the cell in which they are expressed. "Secreted" proteins also include without limitation proteins which are transported across the membrane of the endoplasmic reticulum. As used herein, a "mature protein" is the polypeptide fragment generated after the cleavage of the signal peptide.

The term "full coding sequence" or open reading frame (ORF) of a GENSET gene, when used herein, refers to the complete coding sequence of said gene. In the case of a secreted protein, the full coding sequence comprises the coding sequence for the signal peptide and the coding sequence for the mature polypeptide. Accordingly, the term "full-length polypeptide" refers to the complete polypeptide encoded by said GENSET gene and in the case of a secreted protein it comprises both the signal peptide and the mature polypeptide. The positions of the full length polypeptides and, in the case of secreted proteins, of signal peptides and mature polypeptides are given in the appended sequence listing.

The term "GENSET biological activity" is intended for polypeptides exhibiting an activity similar, but not necessarily identical, to an activity of the GENSET polypeptide of the invention.

The GENSET biological activity of a given polypeptide may be assessed using a suitable biological assay well known to those skilled in the art such as the one(s) described herein. In contrast, the term "biological activity" refers to any activity that a polypeptide of the invention may have.

The term "corresponding mRNA" refers to the mRNA which was the template for the
5 cDNA synthesis which produced a cDNA of the present invention.

The term "corresponding genomic DNA" refers to the genomic DNA which encodes mRNA which includes the sequence of one of the strands of the cDNA in which thymidine residues in the sequence of the cDNA are replaced by uracil residues in the mRNA.

The term "deposited clone pool" is used herein to refer to the pool of clones entitled
10 GENSET.071PRF deposited in ATCC with the accession number PTA-1218 on January, 21, 2000.

The term "heterologous", when used herein, is intended to designate any polynucleotide or polypeptide other than the GENSET polynucleotide or polypeptide respectively.

The term "isolated" requires that the material be removed from its original environment (e. g., the natural environment if it is naturally occurring). For example, a naturally-occurring
15 polynucleotide or polypeptide present in a living animal is not isolated, but the same polynucleotide or DNA or polypeptide, separated from some or all of the coexisting materials in the natural system, is isolated. Such polynucleotide could be part of a vector and/or such polynucleotide or polypeptide could be part of a composition, and still be isolated in that the vector or composition is not part of its natural environment. For example, a naturally-occurring polynucleotide present in a living
20 animal is not isolated, but the same polynucleotide, separated from some or all of the coexisting materials in the natural system, is isolated. Specifically excluded from the definition of "isolated" are: naturally-occurring chromosomes (such as chromosome spreads), artificial chromosome libraries, genomic libraries, and cDNA libraries that exist either as an *in vitro* nucleic acid preparation or as a transfected/transformed host cell preparation, wherein the host cells are either an
25 *in vitro* heterogeneous preparation or plated as a heterogeneous population of single colonies. Also specifically excluded are the above libraries wherein a specified polynucleotide makes up less than 5% of the number of nucleic acid inserts in the vector molecules. Further specifically excluded are whole cell genomic DNA or whole cell RNA preparations (including said whole cell preparations which are mechanically sheared or enzymatically digested). Further specifically excluded are the
30 above whole cell preparations as either an *in vitro* preparation or as a heterogeneous mixture separated by electrophoresis (including blot transfers of the same) wherein the polynucleotide of the invention has not further been separated from the heterologous polynucleotides in the electrophoresis medium (e.g., further separating by excising a single band from a heterogeneous band population in an agarose gel or nylon blot).

35 The term "purified" does not require absolute purity; rather, it is intended as a relative definition. Purification of starting material or natural material to at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is expressly

contemplated. As an example, purification from 0.1 % concentration to 10 % concentration is two orders of magnitude. To illustrate, individual cDNA clones isolated from a cDNA library have been conventionally purified to electrophoretic homogeneity. The sequences obtained from these clones could not be obtained directly either from the library or from total human DNA. The cDNA clones
5 are not naturally occurring as such, but rather are obtained via manipulation of a partially purified naturally occurring substance (messenger RNA). The conversion of mRNA into a cDNA library involves the creation of a synthetic substance (cDNA) and pure individual cDNA clones can be isolated from the synthetic library by clonal selection. Thus, creating a cDNA library from messenger RNA and subsequently isolating individual clones from that library results in an
10 approximately 10^4 - 10^6 fold purification of the native message.

The term "purified" is further used herein to describe a polypeptide or polynucleotide of the invention which has been separated from other compounds including, but not limited to, polypeptides or polynucleotides, carbohydrates, lipids, etc. The term "purified" may be used to specify the separation of monomeric polypeptides of the invention from oligomeric forms such as
15 homo- or hetero- dimers, trimers, etc. The term "purified" may also be used to specify the separation of covalently closed polynucleotides from linear polynucleotides. A polynucleotide is substantially pure when at least about 50%, preferably 60 to 75% of a sample exhibits a single polynucleotide sequence and conformation (linear versus covalently close). A substantially pure polypeptide or polynucleotide typically comprises about 50%, preferably 60 to 90% weight/weight
20 of a polypeptide or polynucleotide sample, respectively, more usually about 95%, and preferably is over about 99% pure. Polypeptide and polynucleotide purity, or homogeneity, is indicated by a number of means well known in the art, such as agarose or polyacrylamide gel electrophoresis of a sample, followed by visualizing a single band upon staining the gel. For certain purposes higher resolution can be provided by using HPLC or other means well known in the art. As an alternative
25 embodiment, purification of the polypeptides and polynucleotides of the present invention may be expressed as "at least" a percent purity relative to heterologous polypeptides and polynucleotides (DNA, RNA or both). As a preferred embodiment, the polypeptides and polynucleotides of the present invention are at least; 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 96%, 96%, 98%, 99%, or 100% pure relative to heterologous polypeptides and polynucleotides, respectively. As a
30 further preferred embodiment the polypeptides and polynucleotides have a purity ranging from any number, to the thousandth position, between 90% and 100% (e.g., a polypeptide or polynucleotide at least 99.995% pure) relative to either heterologous polypeptides or polynucleotides, respectively, or as a weight/weight ratio relative to all compounds and molecules other than those existing in the carrier. Each number representing a percent purity, to the thousandth position, may be claimed as
35 individual species of purity.

As used interchangeably herein, the terms "nucleic acid molecule(s)", "oligonucleotide(s)", and "polynucleotide(s)" include RNA or DNA (either single or double stranded, coding,

complementary or antisense), or RNA/DNA hybrid sequences of more than one nucleotide in either single chain or duplex form (although each of the above species may be particularly specified). The term “nucleotide” is used herein as an adjective to describe molecules comprising RNA, DNA, or RNA/DNA hybrid sequences of any length in single-stranded or duplex form. More precisely, the expression “nucleotide sequence” encompasses the nucleic material itself and is thus not restricted to the sequence information (i.e. the succession of letters chosen among the four base letters) that biochemically characterizes a specific DNA or RNA molecule. The term “nucleotide” is also used herein as a noun to refer to individual nucleotides or varieties of nucleotides, meaning a molecule, or individual unit in a larger nucleic acid molecule, comprising a purine or pyrimidine, a ribose or deoxyribose sugar moiety, and a phosphate group, or phosphodiester linkage in the case of nucleotides within an oligonucleotide or polynucleotide. The term “nucleotide” is also used herein to encompass “modified nucleotides” which comprise at least one modifications such as (a) an alternative linking group, (b) an analogous form of purine, (c) an analogous form of pyrimidine, or (d) an analogous sugar. For examples of analogous linking groups, purine, pyrimidines, and sugars see for example PCT publication No. WO 95/04064, which disclosure is hereby incorporated by reference in its entirety. Preferred modifications of the present invention include, but are not limited to, 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xantine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v) ybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid, 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, and 2,6-diaminopurine. The polynucleotide sequences of the invention may be prepared by any known method, including synthetic, recombinant, *ex vivo* generation, or a combination thereof, as well as utilizing any purification methods known in the art. Methylenemethylimino linked oligonucleosides as well as mixed backbone compounds having, may be prepared as described in U.S. Pat. Nos. 5,378,825; 5,386,023; 5,489,677; 5,602,240; and 5,610,289, which disclosures are hereby incorporated by reference in their entirety. Formacetal and thioformacetal linked oligonucleosides may be prepared as described in U.S. Pat. Nos. 5,264,562 and 5,264,564, which disclosures are hereby incorporated by reference in their entirety. Ethylene oxide linked oligonucleosides may be prepared as described in U.S. Pat. No. 5,223,618, which disclosure is hereby incorporated by reference in its entirety. Phosphinate oligonucleotides may be prepared as described in U.S. Pat. No. 5,508,270, which disclosure is hereby incorporated by reference in its entirety. Alkyl phosphonate

oligonucleotides may be prepared as described in U.S. Pat. No. 4,469,863, which disclosure is hereby incorporated by reference in its entirety. 3'-Deoxy-3'-methylene phosphonate oligonucleotides may be prepared as described in U.S. Pat. Nos. 5,610,289 or 5,625,050 which disclosures are hereby incorporated by reference in their entireties. Phosphoramidite oligonucleotides may be prepared as described in U.S. Pat. No. 5,256,775 or U.S. Pat. No. 5,366,878 which disclosures are hereby incorporated by reference in their entireties. Alkylphosphonothioate oligonucleotides may be prepared as described in published PCT applications WO 94/17093 and WO 94/02499 which disclosures are hereby incorporated by reference in their entireties. 3'-Deoxy-3'-amino phosphoramidate oligonucleotides may be prepared as described in U.S. Pat. No. 5,476,925, which disclosure is hereby incorporated by reference in its entirety. Phosphotriester oligonucleotides may be prepared as described in U.S. Pat. No. 5,023,243, which disclosure is hereby incorporated by reference in its entirety. Borano phosphate oligonucleotides may be prepared as described in U.S. Pat. Nos. 5,130,302 and 5,177,198 which disclosures are hereby incorporated by reference in their entireties.

15 The term "upstream" is used herein to refer to a location which is toward the 5' end of the polynucleotide from a specific reference point.

The terms "base paired" and "Watson & Crick base paired" are used interchangeably herein to refer to nucleotides which can be hydrogen bonded to one another by virtue of their sequence identities in a manner like that found in double-helical DNA with thymine or uracil residues linked to adenine residues by two hydrogen bonds and cytosine and guanine residues linked by three hydrogen bonds (See Stryer, 1995, which disclosure is hereby incorporated by reference in its entirety).

The terms "complementary" or "complement thereof" are used herein to refer to the sequences of polynucleotides which is capable of forming Watson & Crick base pairing with another specified polynucleotide throughout the entirety of the complementary region. For the purpose of the present invention, a first polynucleotide is deemed to be complementary to a second polynucleotide when each base in the first polynucleotide is paired with its complementary base. Complementary bases are, generally, A and T (or A and U), or C and G. "Complement" is used herein as a synonym from "complementary polynucleotide", "complementary nucleic acid" and "complementary nucleotide sequence". These terms are applied to pairs of polynucleotides based solely upon their sequences and not any particular set of conditions under which the two polynucleotides would actually bind. Unless otherwise stated, all complementary polynucleotides are fully complementary on the whole length of the considered polynucleotide.

The terms "polypeptide" and "protein", used interchangeably herein, refer to a polymer of amino acids without regard to the length of the polymer; thus, peptides, oligopeptides, and proteins are included within the definition of polypeptide. This term also does not specify or exclude chemical or post-expression modifications of the polypeptides of the invention, although chemical

or post-expression modifications of these polypeptides may be included or excluded as specific embodiments. Therefore, for example, modifications to polypeptides that include the covalent attachment of glycosyl groups, acetyl groups, phosphate groups, lipid groups and the like are expressly encompassed by the term polypeptide. Further, polypeptides with these modifications may be specified as individual species to be included or excluded from the present invention. The natural or other chemical modifications, such as those listed in examples above can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be branched, for example, as a result of ubiquitination, and they may be cyclic, with or without branching. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, pegylation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. (See, for instance Creighton (1993); Seifter *et al.*, (1990); Rattan *et al.*, (1992)). Also included within the definition are polypeptides which contain one or more analogs of an amino acid (including, for example, non-naturally occurring amino acids, amino acids which only occur naturally in an unrelated biological system, modified amino acids from mammalian systems, etc...), polypeptides with substituted linkages, as well as other modifications known in the art, both naturally occurring and non-naturally occurring.

As used herein, the terms "recombinant polynucleotide" and "polynucleotide construct" are used interchangeably to refer to linear or circular, purified or isolated polynucleotides that have been artificially designed and which comprise at least two nucleotide sequences that are not found as contiguous nucleotide sequences in their initial natural environment. In particular, this terms mean that the polynucleotide or cDNA is adjacent to "backbone" nucleic acid to which it is not adjacent in its natural environment. Additionally, to be "enriched" the cDNAs will represent 5% or more of the number of nucleic acid inserts in a population of nucleic acid backbone molecules. Backbone molecules according to the present invention include nucleic acids such as expression vectors, self-replicating nucleic acids, viruses, integrating nucleic acids, and other vectors or nucleic acids used to maintain or manipulate a nucleic acid insert of interest. Preferably, the enriched cDNAs represent 15% or more of the number of nucleic acid inserts in the population of recombinant backbone molecules. More preferably, the enriched cDNAs represent 50% or more of

the number of nucleic acid inserts in the population of recombinant backbone molecules. In a highly preferred embodiment, the enriched cDNAs represent 90% or more (including any number between 90 and 100%, to the thousandth position, e.g., 99.5%) # of the number of nucleic acid inserts in the population of recombinant backbone molecules.

- 5 The term “recombinant polypeptide” is used herein to refer to polypeptides that have been artificially designed and which comprise at least two polypeptide sequences that are not found as contiguous polypeptide sequences in their initial natural environment, or to refer to polypeptides which have been expressed from a recombinant polynucleotide.

As used herein, the term “operably linked” refers to a linkage of polynucleotide elements in a functional relationship. A sequence which is “operably linked” to a regulatory sequence such as a promoter means that said regulatory element is in the correct location and orientation in relation to the nucleic acid to control RNA polymerase initiation and expression of the nucleic acid of interest. For instance, a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the coding sequence.

- 15 As used herein, the term “non-human animal” refers to any non-human animal, including insects, birds, rodents and more usually mammals. Preferred non-human animals include: primates; farm animals such as swine, goats, sheep, donkeys, cattle, horses, chickens, rabbits; and rodents, preferably rats or mice. As used herein, the term “animal” is used to refer to any species in the animal kingdom, preferably vertebrates, including birds and fish, and more preferable a mammal.
- 20 Both the terms “animal” and “mammal” expressly embrace human subjects unless preceded with the term “non-human”.

The term “domain” refers to an amino acid fragment with specific biological properties. This term encompasses all known structural and linear biological motifs. Examples of such motifs include but are not limited to leucine zippers, helix-turn-helix motifs, glycosylation sites, ubiquitination sites, alpha helices, and beta sheets, signal peptides which direct the secretion of proteins, sites for post-translational modification, enzymatic active sites, substrate binding sites, and enzymatic cleavage sites.

- 25 Although they have distinct meanings, the terms “comprising”, “consisting of” and “consisting essentially of” may be interchanged for one another throughout the instant application.
- 30 The term “having” has the same meaning as “comprising” and may be replaced with either the term “consisting of” or “consisting essentially of”.

An “amplification product” refers to a product of any amplification reaction, e.g. PCR, RT-PCR, LCR, etc.

- A “modulator” of a protein or other compound refers to any agent that has a functional effect on the protein, including physical binding to the protein, alterations of the quantity or quality of expression of the protein, altering any measurable or detectable activity, property, or behavior of the protein, or in any way interacts with the protein or compound.
- 35

“A test compound” can be any molecule that is evaluated for its ability to modulate a protein or other compound.

Unless otherwise specified in the application, nucleotides and amino acids of polynucleotides and polypeptides respectively of the present invention are contiguous and not interrupted by heterologous sequences.

Identity Between Nucleic Acids Or Polypeptides

The terms “percentage of sequence identity” and “percentage homology” are used interchangeably herein to refer to comparisons among polynucleotides and polypeptides, and are determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide or polypeptide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity. Homology is evaluated using any of the variety of sequence comparison algorithms and programs known in the art. Such algorithms and programs include, but are by no means limited to, TBLASTN, BLASTP, FASTA, TFASTA, CLUSTALW, FASTDB (Pearson and Lipman, 1988; Altschul *et al.*, 1990; Thompson *et al.*, 1994; Higgins *et al.*, 1996; Altschul *et al.*, 1990; Altschul *et al.*, 1993; Brutlag *et al.*, 1990), the disclosures of which are incorporated by reference in their entireties.

In a particularly preferred embodiment, protein and nucleic acid sequence homologies are evaluated using the Basic Local Alignment Search Tool (“BLAST”) which is well known in the art (see, e.g., Karlin and Altschul, 1990; Altschul *et al.*, 1990, 1993, 1997), the disclosures of which are incorporated by reference in their entireties. In particular, five specific BLAST programs are used to perform the following task:

- (1) BLASTP and BLAST3 compare an amino acid query sequence against a protein sequence database;
- (2) BLASTN compares a nucleotide query sequence against a nucleotide sequence database;
- (3) BLASTX compares the six-frame conceptual translation products of a query nucleotide sequence (both strands) against a protein sequence database;
- (4) TBLASTN compares a query protein sequence against a nucleotide sequence database translated in all six reading frames (both strands); and
- (5) TBLASTX compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database.

The BLAST programs identify homologous sequences by identifying similar segments, which are referred to herein as "high-scoring segment pairs," between a query amino or nucleic acid sequence and a test sequence which is preferably obtained from a protein or nucleic acid sequence database. High-scoring segment pairs are preferably identified (i.e., aligned) by means of a scoring matrix, many of which are known in the art. Preferably, the scoring matrix used is the BLOSUM62 matrix (Gonnet *et al.*, 1992; Henikoff and Henikoff, 1993), the disclosures of which are incorporated by reference in their entireties. Less preferably, the PAM or PAM250 matrices may also be used (see, e.g., Schwartz and Dayhoff, eds., 1978), the disclosure of which is incorporated by reference in its entirety. The BLAST programs evaluate the statistical significance of all high-scoring segment pairs identified, and preferably selects those segments which satisfy a user-specified threshold of significance, such as a user-specified percent homology. Preferably, the statistical significance of a high-scoring segment pair is evaluated using the statistical significance formula of Karlin (see, e.g., Karlin and Altschul, 1990), the disclosure of which is incorporated by reference in its entirety. The BLAST programs may be used with the default parameters or with modified parameters provided by the user.

Another preferred method for determining the best overall match between a query nucleotide sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag *et al.* (1990), the disclosure of which is incorporated by reference in its entirety. In a sequence alignment the query and subject sequences are both DNA sequences. An RNA sequence can be compared by first converting U's to T's. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB alignment of DNA sequences to calculate percent identity are: Matrix=Unitary, k-tuple=4, Mismatch Penalty= 1, Joining Penalty=30, Randomization Group Length=0, Cutoff Score= 1, Gap Penalty=5, Gap Size Penalty 0.05, Window Size=500 or the length of the subject nucleotide sequence, whichever is shorter. If the subject sequence is shorter than the query sequence because of 5' or 3' deletions, not because of internal deletions, a manual correction must be made to the results. This is because the FASTDB program does not account for 5' and 3' truncations of the subject sequence when calculating percent identity. For subject sequences truncated at the 5' or 3'ends, relative to the query sequence, the percent identity is corrected by calculating the number of bases of the query sequence that are 5' and 3' of the subject sequence, which are not matched/aligned, as a percent of the total bases of the query sequence. Whether a nucleotide is matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This corrected score is what is used for the purposes of the present invention. Only nucleotides outside the 5' and 3' nucleotides of the subject sequence, as displayed by the FASTDB alignment, which are not matched/aligned with the query sequence, are calculated for the

purposes of manually adjusting the percent identity score. For example, a 90 nucleotide subject sequence is aligned to a 100 nucleotide query sequence to determine percent identity. The deletions occur at the 5' end of the subject sequence and therefore, the FASTDB alignment does not show a matched/alignment of the first 10 nucleotides at 5' end. The 10 unpaired nucleotides represent 10% of the sequence (number of nucleotides at the 5' and 3' ends not matched/total number of nucleotides in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 nucleotides were perfectly matched the final percent identity would be 90%. In another example, a 90 nucleotide subject sequence is compared with a 100 nucleotide query sequence. This time the deletions are internal deletions so that there are no nucleotides on the 5' or 3' of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only nucleotides 5' and 3' of the subject sequence which are not matched/aligned with the query sequence are manually corrected. No other manual corrections are made for the purposes of the present invention.

Another preferred method for determining the best overall match between a query amino acid sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag *et al.* (1990). In a sequence alignment the query and subject sequences are both amino acid sequences. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB amino acid alignment are: Matrix=PAM 0, k-tuple=2, Mismatch Penalty= 1, Joining Penalty=20, Randomization Group25Length=0, Cutoff Score= 1, Window Size=sequence length, Gap Penalty=5, Gap Size Penalty=0.05, Window Size=500 or the length of the subject amino acid sequence, whichever is shorter. If the subject sequence is shorter than the query sequence due to N-or C-terminal deletions, not because of internal deletions, the results, in percent identity, must be manually corrected. This is because the FASTDB program does not account for N- and C-terminal truncations of the subject sequence when calculating global percent identity. For subject sequences truncated at the N- and C-termini, relative to the query sequence, the percent identity is corrected by calculating the number of residues of the query sequence that are N- and C- terminal of the subject sequence, which are not matched/aligned with a corresponding subject residue, as a percent of the total bases of the query sequence. Whether a residue is matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This final percent identity score is what is used for the purposes of the present invention. Only residues to the N- and C-termini of the subject sequence, which are not matched/aligned with the query sequence, are considered for the purposes of manually adjusting the percent identity score. That is, only query amino acid residues outside the farthest N- and C-terminal residues of the subject sequence. For example, a 90 amino

acid residue subject sequence is aligned with a 100-residue query sequence to determine percent identity. The deletion occurs at the N-terminus of the subject sequence and therefore, the FASTDB alignment does not match/align with the first residues at the N-terminus. The 10 unpaired residues represent 10% of the sequence (number of residues at the N- and C- termini not matched/total number of residues in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 residues were perfectly matched the final percent identity would be 90%. In another example, a 90-residue subject sequence is compared with a 100-residue query sequence. This time the deletions are internal so there are no residues at the N- or C-termini of the subject sequence, which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only residue positions outside the N- and C-terminal ends of the subject sequence, as displayed in the FASTDB alignment, which are not matched/aligned with the query sequence are manually corrected. No other manual corrections are made for the purposes of the present invention.

The term “percentage of sequence similarity” refers to comparisons between polypeptide sequences and is determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polypeptide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which an identical or equivalent amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence similarity. Similarity is evaluated using any of the variety of sequence comparison algorithms and programs known in the art, including those described above in this section. Equivalent amino acid residues are defined herein in the “Mutated polypeptides” section.

POLYNUCLEOTIDES OF THE INVENTION

The present invention concerns GENSET genomic and cDNA sequences. The present invention encompasses GENSET genes, polynucleotides comprising GENSET genomic and cDNA sequences, as well as fragments and variants thereof. These polynucleotides may be purified, isolated, or recombinant.

Also encompassed by the present invention are allelic variants, orthologs, splice variants, and/or species homologues of the GENSET genes. Procedures known in the art can be used to obtain full-length genes and cDNAs, allelic variants, splice variants, full-length coding portions, orthologs, and/or species homologues of genes and cDNAs corresponding to a nucleotide sequence selected from the group consisting of sequences of SEQ ID Nos: 1-241 and sequences of clone inserts of the deposited clone pool, using information from the sequences disclosed herein or the

clone pool deposited with the ATCC. For example, allelic variants, orthologs and/or species homologues may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source for allelic variants and/or the desired homologue using any technique known to those skilled in the art including those described
5 into the section entitled "To find similar sequences".

In a specific embodiment, the polynucleotides of the invention are at least 15, 30, 50, 100, 125, 500, or 1000 continuous nucleotides. In another embodiment, the polynucleotides are less than or equal to 300kb, 200kb, 100kb, 50kb, 10kb, 7.5kb, 5kb, 2.5kb, 2kb, 1.5kb, or 1kb in length. In a further embodiment, polynucleotides of the invention comprise a portion of the coding sequences,
10 as disclosed herein, but do not comprise all or a portion of any intron. In another embodiment, the polynucleotides comprising coding sequences do not contain coding sequences of a genomic flanking gene (i.e., 5' or 3' to the gene of interest in the genome). In other embodiments, the polynucleotides of the invention do not contain the coding sequence of more than 1000, 500, 250, 100, 75, 50, 25, 20, 15, 10, 5, 4, 3, 2, or 1 naturally occurring genomic flanking gene(s).

15 Deposited clone pool of the invention

Expression of GENSET genes has been shown to lead to the production of at least one mRNA species per GENSET gene, which cDNA sequence is set forth in the appended sequence listing as SEQ ID Nos: 1-241. The cDNAs (SEQ ID Nos: 1-241) corresponding to these GENSET mRNA species were cloned in the vector pBluescriptII SK⁻ (Stratagene) or one of its derivative
20 called pPT (see figure 1). Cells containing the cloned cDNAs of the present invention are maintained in permanent deposit by the inventors at Genset, S.A., 24 Rue Royale, 75008 Paris, France. Table I provides the applicant's internal designation number (column entitled "Internal designation") assigned to each sequence identification number of SEQ ID Nos: 1-482 (column entitled "Seq Id No") and indicates whether the sequence is a nucleic acid sequence or a
25 polypeptide sequence (column entitled "Type"), and in which vector the cDNA was cloned (column entitled "Vector").

Each cDNA can be removed from the Bluescript vector in which it was inserted by performing a NotI Pst I double digestion to produce the appropriate fragment for each clone provided the cDNA sequence of interest does not contain this restriction site within its sequence.
30 The preferable sites for cDNA removal for those clones inserted into pPT are MunI and HindIII, the sites used for cloning provided the cDNA sequence of interest does not contain this restriction site within its sequence. Alternatively, other restriction enzymes of the multicloning site of the vector may be used to recover the desired insert as indicated by the manufacturer or in figure 1.

Pool of cells containing the cDNAs of the invention, from which the cells containing a
35 particular polynucleotide is obtainable, were also deposited with the American Tissue Culture Collection (ATCC), 10801 University Boulevard, Manassas, VA 20110-2209, United States. Each

cDNA clone has been transfected into separate bacterial cells (E-coli) for these composite deposits. In particular, cells containing the sequences of SEQ ID Nos: 1-241 were deposited on January, 21, 2000 in the pool having ATCC Accession No. PTA-1218 and designated GENSET.071PRF.

Bacterial cells containing a particular clone can be obtained from the composite deposit as follows:

An oligonucleotide probe or probes should be designed to the sequence that is known for that particular clone. This sequence can be derived from the sequences provided herein, or from a combination of those sequences. The design of the oligonucleotide probe should preferably follow these parameters:

(a) It should be designed to an area of the sequence which has the fewest ambiguous bases ("N's"), if any;

(b) Preferably, the probe is designed to have a T_m of approximately 80 degree Celsius (assuming 2 degrees for each A or T and 4 degrees for each G or C). However, probes having melting temperatures between 40 degree Celsius and 80 degree Celsius may also be used provided that specificity is not lost.

The oligonucleotide should preferably be labeled with gamma[32 P]ATP (specific activity 6000 Ci/mmol) and T4 polynucleotide kinase using commonly employed techniques for labeling oligonucleotides. Other labeling techniques can also be used. Unincorporated label should preferably be removed by gel filtration chromatography or other established methods. The amount of radioactivity incorporated into the probe should be quantified by measurement in a scintillation counter. Preferably, specific activity of the resulting probe should be approximately 4×10^6 dpm/pmol.

The bacterial culture containing the pool of full-length clones should preferably be thawed and 100 μ l of the stock used to inoculate a sterile culture flask containing 25 ml of sterile L-broth containing ampicillin at 100 μ g/ml. The culture should preferably be grown to saturation at 37 degree Celsius, and the saturated culture should preferably be diluted in fresh L-broth. Aliquots of these dilutions should preferably be plated to determine the dilution and volume which will yield approximately 5000 distinct and well-separated colonies on solid bacteriological media containing L-broth containing ampicillin at 100 μ g/ml and agar at 1.5% in a 150 mm petri dish when grown overnight at 37 degree Celsius. Other known methods of obtaining distinct, well-separated colonies can also be employed.

Standard colony hybridization procedures should then be used to transfer the colonies to nitrocellulose filters and lyse, denature and bake them.

The filter is then preferably incubated at 65 degree Celsius for 1 hour with gentle agitation in 6X SSC (20X stock is 175.3 g NaCl/liter, 88.2 g Na citrate/liter, adjusted to pH 7.0 with NaOH) containing 0.5% SDS, 100 μ g/ml of yeast RNA, and 10 mM EDTA (approximately 10 ml per 150 mm filter). Preferably, the probe is then added to the hybridization mix at a concentration greater

than or equal to 1×10^6 dpm/ml. The filter is then preferably incubated at 65 degree Celsius with gentle agitation overnight. The filter is then preferably washed in 500 ml of 2X SSC/0.1% SDS at room temperature with gentle shaking for 15 minutes. A third wash with 0.1X SSC/0.5% SDS at 65 degree Celsius for 30 minutes to 1 hour is optional. The filter is then preferably dried and subjected to autoradiography for sufficient time to visualize the positives on the X-ray film. Other known hybridization methods can also be employed.

The positive colonies are picked, grown in culture, and plasmid DNA isolated using standard procedures. The clones can then be verified by restriction analysis, hybridization analysis, or DNA sequencing. The plasmid DNA obtained using these procedures may then be manipulated using standard cloning techniques familiar to those skilled in the art.

Alternatively, to recover cDNA inserts from the pool of bacteria, a PCR can be performed on plasmid DNA isolated using standard procedures and primers designed at both ends of the cDNA insertion, including primers designed in the multicloning site of the vector. For example, a PCR reaction may be conducted using universal primers designed by the plasmid provider or using primers which are specific to the cDNA of interest. In the case of Bluescript SK(-), a PCR reaction may be conducted using a primer having the sequence GGAAACAGCTATGACCA and a primer having the sequence GTAAAACGACGGCCAGT. This will produce a DNA fragment including a piece of the multiple cloning site and the cDNA insert. If a specific cDNA of interest is to be recovered, primers may be designed in order to be specific for the 5' end and the 3' end of this cDNA using sequence information available from the appended sequence listing. The PCR product which corresponds to the cDNA of interest can then be manipulated using standard cloning techniques familiar to those skilled in the art.

Therefore, an object of the invention is an isolated, purified, or recombinant polynucleotide comprising a nucleotide sequence selected from the group consisting of cDNA inserts of the deposited clone pool. Moreover, preferred polynucleotides of the invention include purified, isolated, or recombinant GENSET cDNAs consisting of, consisting essentially of, or comprising a nucleotide sequence selected from the group consisting of cDNA inserts of the deposited clone pool.

The polynucleotides of SEQ ID NOs: 1-141 may be interchanged with the corresponding polynucleotides encoded by the human cDNA of the clones inserts of the deposited clone pool. The polypeptides of SEQ ID NOs: 242-482 may be interchanged with the corresponding polypeptides encoded by the human cDNA of the clones inserts of the deposited clone pool. The correspondance between the polynucleotides of SEQ ID Nos: 1-141, the polypeptides of SEQ ID NOs: 242-482 and clones inserts of the deposited clone pool is given in Table I..

cDNA sequences of the invention

Another object of the invention is a purified, isolated, or recombinant polynucleotide comprising a nucleotide sequence selected from the group consisting of SEQ ID Nos: 1-241, complementary sequences thereto, and fragments thereof. Moreover, preferred

- 5 polynucleotides of the invention include purified, isolated, or recombinant GENSET cDNAs consisting of, consisting essentially of, or comprising a sequence selected from the group consisting of SEQ ID Nos: 1-241.

Polynucleotides GENSET sequences of SEQ ID Nos: 1-241 were then searched for open reading frames able to encode polypeptides. The GENSET ORFs were also searched to identify
10 potential signal sequence motifs using slight modifications of the procedures disclosed in Von Heijne, *Nucleic Acids Res.* 14:4683-4690, 1986, as described in PCT publication WO 00/37491, the entire disclosures of which are incorporated herein by reference. The GENSET cDNAs of SEQ ID Nos: 1-31 and 33-143 encoding polypeptides of SEQ ID Nos: 242-272 and 274-384 were thus found as containing such signal sequences.

15 Structural parameters of each of the cDNA of the present invention are described in Table II. Namely, Table II provides, for each cDNA of SEQ ID Nos: 1-241 referred to by its sequence identification number (column entitled "Seq Id No"), the locations of the first and last nucleotides of the coding sequences (listed under the heading "Full Coding Sequence"), and, if applicable, the locations of the signal sequence and the sequence encoding the mature polypeptide in the case of
20 secreted proteins (SEQ ID Nos: 1-31 and 33-143) listed under the headings "Signal Sequence" and "Coding Sequence for the mature Protein" respectively, the locations of the first and last nucleotides of the polyA signals (listed under the heading "Poly A Signal") and the locations of the first and last nucleotides of the polyA sites (listed under the heading "Poly A Site").

Accordingly, the full coding sequence (CDS) or open reading frame (ORF) of each cDNA
25 of the invention refers to the nucleotide sequence beginning with the first nucleotide of the start codon and ending with the last nucleotide of the stop codon (see column entitled "Full coding sequence" of Table II for sequences of Seq Id Nos: 1-241). Similarly, the signal sequence of each cDNA of the invention refers to the nucleotide sequence beginning with the first nucleotide of the start codon and ending with the last nucleotide of the codon encoding the signal peptide (see
30 column entitled "Signal sequence" of Table II for sequences of Seq Id Nos: 1-31 and 33-143) and the coding sequence for the mature polypeptide of each cDNA of the invention refers to the nucleotide sequence beginning with the first nucleotide of the first codon encoding and ending with the last nucleotide of the stop codon (see column entitled "Coding sequence for mature protein" of Table II for sequences of Seq Id Nos: 1-31 and 33-143). Similarly, the 5'untranslated region (or
35 5'UTR) of each cDNA of the invention refers to the nucleotide sequence starting at nucleotide 1 and ending at the nucleotide immediately 5' to the first nucleotide of the start codon. The 3'untranslated region (or 3'UTR) of each cDNA of the invention refers to the nucleotide sequence

starting at the nucleotide immediately 3' to the last nucleotide of the stop codon and ending at the last nucleotide of the cDNA.

Untranslated regions

In addition, the invention concerns a purified, isolated, and recombinant nucleic acid comprising a nucleotide sequence selected from the group consisting of the 5'UTRs of sequences of SEQ ID Nos: 1-241 and sequences of clone inserts of the deposited clone pool, sequences complementary thereto, and allelic variants thereof. The invention also concerns a purified, isolated, and recombinant nucleic acid comprising a nucleotide sequence selected from the group consisting of the 3'UTRs of sequences of SEQ ID Nos: 1-241 and sequences of clone inserts of the deposited clone pool, sequences complementary thereto, and allelic variants thereof.

These polynucleotides may be used to detect the presence of GENSET mRNA species in a biological sample using either hybridization or RT-PCR techniques well known to those skilled in the art those skilled in the art.

In addition, these polynucleotides may be used as regulatory molecules able to affect the processing and maturation of the polynucleotide including them (either a GENSET polynucleotide or an heterologous polynucleotide), preferably the localization, stability and/or translation of said polynucleotide including them (for a review on UTRs see Decker and Parker, 1995, Derrigo *et al.*, 2000). In particular, 3'UTRs may be used in order to control the stability of heterologous mRNAs in recombinant vectors using any methods known to those skilled in the art including Makrides (1999), US Patents 5,925,56; 5,807,7 and 5,756,264, which disclosures are hereby incorporated by reference in their entireties.

Coding sequences

Another object of the invention is an isolated, purified or recombinant polynucleotide comprising the full coding sequence of a sequence selected from the group consisting of sequences of SEQ ID Nos: 1-241, clone inserts of the deposited clone pool, and variants thereof.

A further object of the invention is an isolated, purified or recombinant polynucleotide encoding a polypeptide comprising a sequence selected from the group consisting of sequences of SEQ ID Nos: 242-482 and allelic variants thereof. Another object of the invention is an isolated, purified or recombinant polynucleotide encoding a polypeptide comprising a sequence selected from the group consisting of polypeptides encoded by cDNA inserts of the deposited clone pool and allelic variants thereof.

In a preferred embodiment, the invention encompasses an isolated, purified or recombinant polynucleotide encoding a polypeptide comprising a sequence selected from the group consisting of the mature proteins of SEQ ID Nos: 242-272 and 274-384. In another preferred embodiment, the invention encompasses an isolated, purified or recombinant polynucleotide encoding a polypeptide

comprising a sequence selected from the group consisting of the signal peptides of SEQ ID Nos: 242-272 and 274-384.

It will be appreciated that should the extent of the full coding sequence differ from that indicated in the appended sequence listing as a result of a sequencing error, reverse transcription or amplification error, mRNA splicing, post-translational modification of the encoded protein, enzymatic cleavage of the encoded protein, or other biological factors, one skilled in the art would be readily able to identify the extent of the full coding sequences in the sequences of SEQ ID Nos: 1-241. Accordingly, the scope of any claims herein relating to nucleic acids containing the full coding sequence of one of SEQ ID Nos: 1-241 is not to be construed as excluding any readily identifiable variations from or equivalents to the full coding sequences described in the appended sequence listing. Similarly, should the extent of the polypeptides differ from those indicated in the appended sequence listing as a result of any of the preceding factors, the scope of claims relating to polypeptides comprising the amino acid sequence of the polypeptides of SEQ ID Nos: 242-482 is not to be construed as excluding any readily identifiable variations from or equivalents to the sequences described in the appended sequence listing.

It will be appreciated that should the extent of the coding sequence of the mature protein differ from that indicated in the appended sequence listing as a result of a sequencing error, reverse transcription or amplification error, mRNA splicing, post-translational modification of the encoded protein, enzymatic cleavage of the encoded protein, or other biological factors, one skilled in the art would be readily able to identify the extent of the coding sequences for the mature protein in the sequences of SEQ ID Nos: 1-31 and 33-143. Accordingly, the scope of any claims herein relating to nucleic acids containing the coding sequence for the mature proteins of one of SEQ ID Nos: 1-31 and 33-143 is not to be construed as excluding any readily identifiable variations from or equivalents to the coding sequences described in the appended sequence listing. Similarly, should the extent of the mature polypeptides differ from those indicated in the appended sequence listing as a result of any of the preceding factors, the scope of claims relating to mature polypeptides comprising the amino acid sequence of the polypeptides of SEQ ID Nos: 242-272 and 274-384 is not to be construed as excluding any readily identifiable variations from or equivalents to the sequences described in the appended sequence listing.

It will be appreciated that should the extent of the coding sequence of the signal peptide differ from that indicated in the appended sequence listing as a result of a sequencing error, reverse transcription or amplification error, mRNA splicing, post-translational modification of the encoded protein, enzymatic cleavage of the encoded protein, or other biological factors, one skilled in the art would be readily able to identify the extent of the coding sequences for the signal peptide in the sequences of SEQ ID Nos: 1-31 and 33-143. Accordingly, the scope of any claims herein relating to nucleic acids containing the signal sequence of one of SEQ ID Nos: 1-31 and 33-143 is not to be construed as excluding any readily identifiable variations from or equivalents to the coding

sequences described in the appended sequence listing. Similarly, should the extent of the signal peptides differ from those indicated in the appended sequence listing as a result of any of the preceding factors, the scope of claims relating to signal peptides comprising the amino acid sequence of the polypeptides of SEQ ID Nos: 242-272 and 274-384 is not to be construed as
5 excluding any readily identifiable variations from or equivalents to the sequences described in the appended sequence listing.

The above disclosed polynucleotides that contains the coding sequence (for the full-length protein or for the mature protein) of the GENSET genes may be expressed in a desired host cell or a desired host organism, when this polynucleotide is placed under the control of suitable expression
10 signals. The expression signals may be either the expression signals contained in the regulatory regions in the GENSET genes of the invention or in contrast the signals may be exogenous regulatory nucleic sequences. Such a polynucleotide, when placed under the suitable expression signals, may also be inserted in a vector for its expression and/or amplification.

Further included in the present invention are polynucleotides encoding the polypeptides of
15 the present invention that are fused in frame to the coding sequences for additional heterologous amino acid sequences. Of special interest are polynucleotides comprising GENSET signal sequences fused to an heterologous polypeptide as described in the section entitled "Secretion vectors". Also included in the present invention are nucleic acids encoding polypeptides of the present invention together with additional, non-coding sequences, including for example, but not
20 limited to non-coding 5' and 3' sequences, vector sequence, sequences used for purification, probing, or priming. For example, heterologous sequences include transcribed, untranslated sequences that may play a role in transcription, and mRNA processing, for example, ribosome binding and stability of mRNA. The heterologous sequences may alternatively comprise additional coding sequences that provide additional functionalities. Thus, a nucleotide sequence encoding a
25 polypeptide may be fused to a tag sequence, such as a sequence encoding a peptide that facilitates purification of the fused polypeptide. In certain preferred embodiments of this aspect of the invention, the tag amino acid sequence is a hexa-histidine peptide, such as the tag provided in a pQE vector (QIAGEN), among others, many of which are commercially available. For instance, hexa-histidine provides for convenient purification of the fusion protein (See Gentz *et al.*, 1989), the
30 disclosure of which is incorporated by reference in its entirety. The "HA" tag is another peptide useful for purification which corresponds to an epitope derived from the influenza hemagglutinin protein (See Wilson *et al.*, 1984), the disclosure of which is incorporated by reference in its entirety. As discussed below other such fusion proteins include the GENSET protein fused to Fc at the N- or C-terminus.

35 Suitable recombinant vectors that contain a polynucleotide such as described herein are disclosed elsewhere in the specification. Expression vectors encoding GENSET polypeptides or fragments thereof are described in the section entitled "Preparation of the polypeptides".

Regulatory sequences of the invention

As mentioned, the genomic sequence of GENSET genes contains regulatory sequences in the non-coding 5'-flanking region and possibly in the non-coding 3'-flanking region that border the GENSET coding regions containing the exons of these genes.

- 5 Polynucleotides derived from GENSET 5' and 3' regulatory regions are useful in order to detect the presence of at least a copy of a genomic nucleotide sequence of the GENSET gene or a fragment thereof in a test sample.

Preferred regulatory sequences

- 10 Polynucleotides carrying the regulatory elements located at the 5' end and at the 3' end of GENSET coding regions may be advantageously used to control the transcriptional and translational activity of a heterologous polynucleotide of interest.

- Thus, the present invention also concerns a purified or isolated nucleic acid comprising a polynucleotide which is selected from the group consisting of the 5' and 3' GENSET regulatory regions, sequences complementary thereto, regulatory active fragments and variants thereof. The
15 invention also pertains to a purified or isolated nucleic acid comprising a polynucleotide having at least 95% nucleotide identity with a polynucleotide selected from the group consisting of GENSET 5' and 3' regulatory regions, advantageously 99 % nucleotide identity, preferably 99.5% nucleotide identity and most preferably 99.8% nucleotide identity with a polynucleotide selected from the group consisting of GENSET 5' and 3' regulatory regions, sequences complementary thereto,
20 variants and regulatory active fragments thereof.

- Another object of the invention consists of purified, isolated or recombinant nucleic acids comprising a polynucleotide that hybridizes, under the stringent hybridization conditions defined herein, with a polynucleotide selected from the group consisting of the nucleotide sequences of GENSET 5'- and 3' regulatory regions, sequences complementary thereto, variants and regulatory
25 active fragments thereof.

Preferred fragments of 5' regulatory regions have a length of about 1500 or 1000 nucleotides, preferably of about 500 nucleotides, more preferably about 400 nucleotides, even more preferably 300 nucleotides and most preferably about 200 nucleotides.

- Preferred fragments of 3' regulatory regions are at least 20, 50, 100, 150, 200, 300 or 400
30 bases in length.

- "Providing" with respect to, e.g. a biological sample, population of cells, etc. indicates that the sample, population of cells, etc. is somehow used in a method or procedure. Significantly, "providing" a biological sample or population of cells does not require that the sample or cells are specifically isolated or obtained for the purposes of the invention, but can instead refer, for
35 example, to the use of a biological sample obtained by another individual, for another purpose.

“Regulatory active” polynucleotide derivatives of the 5' regulatory region are polynucleotides comprising or alternatively consisting of a fragment of said polynucleotide which is functional as a regulatory region for expressing a recombinant polypeptide or a recombinant polynucleotide in a recombinant cell host. It could act either as an enhancer or as a repressor. For the purpose of the invention, a nucleic acid or polynucleotide is “functional” as a regulatory region for expressing a recombinant polypeptide or a recombinant polynucleotide if said regulatory polynucleotide contains nucleotide sequences which contain transcriptional and translational regulatory information, and such sequences are “operably linked” to nucleotide sequences which encode the desired polypeptide or the desired polynucleotide.

10 The regulatory polynucleotides of the invention may be prepared from the nucleotide sequence of GENSET genomic or cDNA sequence, for example, by cleavage using suitable restriction enzymes, or by PCR. The regulatory polynucleotides may also be prepared by digestion of a GENSET gene containing genomic clone by an exonuclease enzyme, such as Bal31 (Wabiko *et al.*, 1986), the disclosure of which is incorporated by reference in its entirety. These regulatory polynucleotides can also be prepared by nucleic acid chemical synthesis, as described elsewhere in the specification.

The regulatory polynucleotides according to the invention may be part of a recombinant expression vector that may be used to express a full coding sequence in a desired host cell or host organism. The recombinant expression vectors according to the invention are described elsewhere in the specification.

Preferred 5'-regulatory polynucleotide of the invention include 5'-UTRs of GENSET cDNAs, or regulatory active fragments or variants thereof. More preferred 5'-regulatory polynucleotides of the invention include sequences selected from the group consisting of 5'-UTRs of sequences of SEQ ID Nos: 1-241, 5'-UTRs of clones inserts of the deposited clone pool, regulatory active fragments and variants thereof.

Preferred 3'-regulatory polynucleotide of the invention include 3'-UTRs of GENSET cDNAs, or regulatory active fragments or variants thereof. More preferred 3'-regulatory polynucleotides of the invention include sequences selected from the group consisting of 3'-UTRs of sequences of SEQ ID Nos: 1-241, 3'-UTRs of clones inserts of the deposited clone pool, regulatory active fragments and variants thereof.

A further object of the invention consists of a purified or isolated nucleic acid comprising:
a) a polynucleotide comprising a 5' regulatory nucleotide sequence selected from the group consisting of:

(i) a nucleotide sequence comprising a polynucleotide of a GENSET 5' regulatory region or a complementary sequence thereto;

(ii) a nucleotide sequence comprising a polynucleotide having at least 95% of nucleotide identity with the nucleotide sequence of a GENSET 5' regulatory region or a complementary sequence thereto;

(iii) a nucleotide sequence comprising a polynucleotide that hybridizes under stringent hybridization conditions with the nucleotide sequence of a GENSET 5' regulatory region or a complementary sequence thereto; and

(iv) a regulatory active fragment or variant of the polynucleotides in (i), (ii) and (iii);

b) a nucleic acid molecule encoding a desired polypeptide or a nucleic acid molecule of interest, said nucleic acid molecule is operably linked to the polynucleotide defined in (a); and

10 c) optionally, a polynucleotide comprising a 3'- regulatory polynucleotide, preferably a 3'- regulatory polynucleotide of a GENSET gene.

In a specific embodiment, the nucleic acid defined above includes the 5'-UTR of a GENSET cDNA, or a regulatory active fragment or variant thereof.

In a second specific embodiment, the nucleic acid defined above includes the 3'-UTR of a
15 GENSET cDNA, or a regulatory active fragment or variant thereof.

The regulatory polynucleotide of the 5' regulatory region, or its regulatory active fragments or variants, is operably linked at the 5'-end of the nucleic acid molecule encoding the desired polypeptide or nucleic acid molecule of interest.

The regulatory polynucleotide of the 3' regulatory region, or its regulatory active fragments
20 or variants, is advantageously operably linked at the 3'-end of the nucleic acid molecule encoding the desired polypeptide or nucleic acid molecule of interest.

The desired polypeptide encoded by the above-described nucleic acid may be of various nature or origin, encompassing proteins of prokaryotic viral or eukaryotic origin. Among the polypeptides expressed under the control of a GENSET regulatory region include bacterial, fungal
25 or viral antigens. Also encompassed are eukaryotic proteins such as intracellular proteins, such as "house keeping" proteins, membrane-bound proteins, such as mitochondrial membrane-bound proteins and cell surface receptors, and secreted proteins such as endogenous mediators such as cytokines. The desired polypeptide may be an heterologous polypeptide or a GENSET protein, especially a protein with an amino acid sequence selected from the group consisting of sequences of
30 SEQ ID Nos: 242-482, fragments and variants thereof.

The desired nucleic acids encoded by the above-described polynucleotide, usually an RNA molecule, may be complementary to a desired coding polynucleotide, for example to a GENSET coding sequence, and thus useful as an antisense polynucleotide. Such a polynucleotide may be included in a recombinant expression vector in order to express the desired polypeptide or the
35 desired nucleic acid in host cell or in a host organism. Suitable recombinant vectors that contain a polynucleotide such as described herein are disclosed elsewhere in the specification. When a

polynucleotide sequence has been recombinantly introduced into a host cell, the cell is said to be "recombinant" for the polynucleotide.

Polynucleotide variants

The invention also relates to variants of the polynucleotides described herein and fragments thereof. "Variants" of polynucleotides, as the term is used herein, are polynucleotides that differ from a reference polynucleotide. Generally, differences are limited so that the nucleotide sequences of the reference and the variant are closely similar overall and, in many regions, identical. The present invention encompasses both allelic variants and degenerate variants.

Examples of variant sequences of polynucleotides of the invention are given in the appended sequence listing. Table III lists the sequence identification number of all similar sequences of the sequence listing, namely variants. All cDNAs referred to by their sequence identification number on a given line of the table are thought to be variants of the same GENSET gene.

Allelic variant

A variant of a polynucleotide may be a naturally occurring variant such as a naturally occurring allelic variant, or it may be a variant that is not known to occur naturally. By an "allelic variant" is intended one of several alternate forms of a gene occupying a given locus on a chromosome of an organism (see Lewin, 1990), the disclosure of which is incorporated by reference in its entirety. Diploid organisms may be homozygous or heterozygous for an allelic form. Non-naturally occurring variants of the polynucleotide may be made by art-known mutagenesis techniques, including those applied to polynucleotides, cells or organisms.

Degenerate variant

In addition to the isolated polynucleotides of the present invention, and fragments thereof, the invention further includes polynucleotides which comprise a sequence substantially different from those described above but which, due to the degeneracy of the genetic code, still encode a GENSET polypeptide of the present invention. These polynucleotide variants are referred to as "degenerate variants" throughout the instant application. That is, all possible polynucleotide sequences that encode the GENSET polypeptides of the present invention are completed. This includes the genetic code and species-specific codon preferences known in the art. Thus, it would be routine for one skilled in the art to generate the degenerate variants described above, for instance, to optimize codon expression for a particular host (e.g., change codons in the human mRNA to those preferred by other mammalian or bacterial host cells).

Nucleotide changes present in a variant polynucleotide may be silent, which means that they do not alter the amino acids encoded by the polynucleotide. However, nucleotide changes may also result in amino acid substitutions, additions, deletions, fusions and truncations in the

polypeptide encoded by the reference sequence. The substitutions, deletions or additions may involve one or more nucleotides. The variants may be altered in coding or non-coding regions or both. Alterations in the coding regions may produce conservative or non-conservative amino acid substitutions, deletions or additions. In the context of the present invention, preferred embodiments
5 are those in which the polynucleotide variants encode polypeptides which retain substantially the same biological properties or activities as the GENSET protein. More preferred polynucleotide variants are those containing conservative substitutions.

Similar polynucleotides

Other embodiments of the present invention is a purified, isolated or recombinant
10 polynucleotide which is at least 90%, 95%, 96%, 97%, 98% or 99% identical to a polynucleotide selected from the group consisting of sequences of SEQ ID Nos: 1-241 and clone inserts of the deposited clone pool. The above polynucleotides are included regardless of whether they encode a polypeptide having a GENSET biological activity. This is because even where a particular nucleic acid molecule does not encode a polypeptide having activity, one of skill in the art would still know
15 how to use the nucleic acid molecule, for instance, as a hybridization probe or primer. Uses of the nucleic acid molecules of the present invention that do not encode a polypeptide having GENSET activity include, inter alia, isolating a GENSET gene or allelic variants thereof from a DNA library, and detecting GENSET mRNA expression in biological samples, suspected of containing GENSET mRNA or DNA by Northern Blot or PCR analysis.

20 The present invention is further directed to polynucleotides having sequences at least 50%. 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98% or 99% identity to a polynucleotide selected from the group consisting of sequences of SEQ ID Nos: 1-241 and clone inserts of the deposited clone pool, where said polynucleotide do, in fact, encode a polypeptide having a GENSET biological activity. Of course, due to the degeneracy of the genetic code, one of ordinary skill in the art will
25 immediately recognize that a large number of the polynucleotides at least 50%. 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, or 99% identical to a polynucleotide selected from the group consisting of sequences of SEQ ID Nos: 1-241 and clone inserts of the deposited clone pool will encode a polypeptide having biological activity. In fact, since degenerate variants of these nucleotide sequences all encode the same polypeptide, this will be clear to the skilled artisan even
30 without performing the above described comparison assay. It will be further recognized in the art that, for such nucleic acid molecules that are not degenerate variants, a reasonable number will also encode a polypeptide having biological activity. This is because the skilled artisan is fully aware of amino acid substitutions that are either less likely or not likely to significantly effect protein function (e.g., replacing one aliphatic amino acid with a second aliphatic amino acid), as further
35 described below. By a polynucleotide having a nucleotide sequence at least, for example, 95% "identical" to a reference nucleotide sequence of the present invention, it is intended that the

nucleotide sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence encoding the GENSET polypeptide. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted, inserted, or substituted with another nucleotide. The query sequence may be an entire sequence selected from the group consisting of sequences of SEQ ID Nos: 1-241 and sequences of clone inserts of the deposited clone pool, or the ORF (open reading frame) of a polynucleotide sequence selected from said group, or any fragment specified as described herein.

10 Hybridizing Polynucleotides

In another aspect, the invention provides an isolated or purified nucleic acid molecule comprising a polynucleotide which hybridizes under stringent hybridization conditions to any polynucleotide of the present invention using any methods known to those skilled in the art including those disclosed herein and in particular in the "To find similar sequences" section. Also contemplated are nucleic acid molecules that hybridize to the polynucleotides of the present invention at lower stringency hybridization conditions, preferably at moderate or low stringency conditions as defined herein. Such hybridizing polynucleotides may be of at least 15, 18, 20, 23, 25, 28, 30, 35, 40, 50, 75, 100, 200, 300, 500, 1000 or 2000 nucleotides in length.

Of particular interest, are the polynucleotides hybridizing to any polynucleotide of the invention and encoding GENSET polypeptides, particularly GENSET polypeptides exhibiting a GENSET biological activity.

Of course, a polynucleotide which hybridizes only to polyA⁺ sequences (such as any 3' terminal polyA⁺ tract of a cDNA shown in the sequence listing), or to a 5' complementary stretch of T (or U) residues, would not be included in the definition of "polynucleotide," since such a polynucleotide would hybridize to any nucleic acid molecule containing a poly (A) stretch or the complement thereof (e.g., practically any double-stranded cDNA clone generated using oligo dT as a primer).

Complementary polynucleotides

The invention further provides isolated nucleic acid molecules having a nucleotide sequence fully complementary to any polynucleotide of the invention. The present invention encompasses a purified, isolated or recombinant polynucleotide having a nucleotide sequence complementary to a sequence selected from the group consisting of sequences of SEQ ID Nos: 1-241, sequences of clone inserts of the deposited clone pool and fragments thereof. Such isolated molecules, particularly DNA molecules, are useful as probes for gene mapping and for identifying GENSET mRNA in a biological sample, for instance, by PCR or Northern blot analysis.

Polynucleotides fragments

The present invention is further directed to polynucleotides encoding portions or fragments of the nucleotide sequences described herein. Uses for the polynucleotide fragments of the present invention include probes, primers, molecular weight markers and for expressing the polypeptide fragments of the present invention. Fragments include portions of polynucleotides selected from the group consisting of a) the sequences of SEQ ID Nos: 1-241, b) the genomic GENSET sequences, c) the polynucleotides encoding a polypeptide selected from the group consisting of the sequences of SEQ ID Nos: 242-482, d) the sequences of clone inserts of the deposited clone pool, and e) the polynucleotides encoding the polypeptides encoded by the clone inserts of the deposited clone pool. Particularly included in the present invention is a purified or isolated polynucleotide comprising at least 8 consecutive bases of a polynucleotide of the present invention. In one aspect of this embodiment, the polynucleotide comprises at least 10, 12, 15, 18, 20, 25, 28, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, 500, 800, 1000, 1500, or 2000 consecutive nucleotides of a polynucleotide of the present invention.

In addition to the above preferred polynucleotide sizes, further preferred sub-genuses of polynucleotides comprise at least 8 nucleotides, wherein "at least 8" is defined as any integer between 8 and the integer representing the 3' most nucleotide position as set forth in the sequence listing or elsewhere herein. Further included as preferred polynucleotides of the present invention are polynucleotide fragments at least 8 nucleotides in length, as described above, that are further specified in terms of their 5' and 3' position. The 5' and 3' positions are represented by the position numbers set forth in the appended sequence listing. For allelic, degenerate and other variants, position 1 is defined as the 5' most nucleotide of the ORF, i.e., the nucleotide "A" of the start codon with the remaining nucleotides numbered consecutively. Therefore, every combination of a 5' and 3' nucleotide position that a polynucleotide fragment of the present invention, at least 8 contiguous nucleotides in length, could occupy on a polynucleotide of the invention is included in the invention as an individual species. The polynucleotide fragments specified by 5' and 3' positions can be immediately envisaged and are therefore not individually listed solely for the purpose of not unnecessarily lengthening the specifications.

It is noted that the above species of polynucleotide fragments of the present invention may alternatively be described by the formula "a to b"; where "a" equals the 5' most nucleotide position and "b" equals the 3' most nucleotide position of the polynucleotide; and further where "a" equals an integer between 1 and the number of nucleotides of the polynucleotide sequence of the present invention minus 8, and where "b" equals an integer between 9 and the number of nucleotides of the polynucleotide sequence of the present invention; and where "a" is an integer smaller than "b" by at least 8.

The present invention also provides for the exclusion of any species of polynucleotide fragments of the present invention specified by 5' and 3' positions or sub-genuses of

polynucleotides specified by size in nucleotides as described above. Any number of fragments specified by 5' and 3' positions or by size in nucleotides, as described above, may be excluded. Specifically excluded from the invention are the fragments described in Table IV. For these cDNAs referred to by their sequence identification numbers, Table IV gives the positions of excluded

5 fragments within these sequences fragments having substantial homology to polyadenylation tails and to repeated sequences including Alu, L1, THE and MER repeats, SSTR sequences or satellite, micro-satellite, and telomeric repeats. Each fragment is represented by a-b where a and b are the start and end positions respectively of a given excluded fragment. Excluded fragments are separated from each other by a coma. As used herein the term "polynucleotide described in Table IV" refers

10 to all polynucleotide fragments defined in Table IV in this manner.

Preferred included and excluded polynucleotide fragments of the invention are also described in Tables Va and Table Vb. For these cDNAs referred to by their sequence identification numbers, Tables Va and Table Vb give the positions of preferred fragments within these sequences (columns entitled "Preferentially included fragments") as well as the positions of preferentially

15 excluded fragments (columns entitled "Preferentially excluded fragments"). Each fragment is represented by a-b where a and b are the start and end positions respectively of a given preferred fragment. Fragments are separated from each other by a coma. As used herein the term "excluded polynucleotide described in Tables Va and Vb" refers to all polynucleotide preferentially excluded as described in Tables Va and Vb. As used herein the term "preferred polynucleotide described in

20 Tables Va and Vb" refers to all preferentially included polynucleotide fragments listed in Tables Va and Table Vb in this manner.

Therefore, the present invention encompasses isolated, purified, or recombinant polynucleotides which consist of, consist essentially of, or comprise a contiguous span of at least 8, 10, 12, 15, 18, 20, 25, 28, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, 500, 1000 or 2000 nucleotides

25 of a sequence selected from the group consisting of the sequences of SEQ ID Nos: 1-241 and sequences fully complementary thereto, to the extent that a contiguous span of these lengths is consistent with the lengths of said selected sequence, wherein said contiguous span comprises at least 1, 2, 3, 5, 10, 15, 18, 20, 25, 28, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400 or 500 nucleotides of a preferred polynucleotide described in Tables Va and Vb, or a sequence complementary thereto.

30 The present invention also encompasses isolated, purified, or recombinant polynucleotides comprising, consisting essentially of, or consisting of a contiguous span of at least 8, 10, 12, 15, 18, 20, 25, 28, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, 500, 1000 or 2000 nucleotides of a polynucleotide selected from the group consisting of the sequences of SEQ ID Nos: 1-241 and sequences fully complementary thereto, wherein said contiguous span comprises a preferred

35 polynucleotide described in Tables Va and Vb, or a sequence complementary thereto, to the extent that a contiguous span of these lengths is consistent with the length of the selected sequence. The present invention also encompasses isolated, purified, or recombinant nucleic acids which comprise,

consist of or consist essentially of a contiguous span of a polynucleotide selected from the group consisting of the sequences of SEQ ID Nos: 1-241 and sequences fully complementary thereto, wherein said contiguous span comprises preferred polynucleotide described in Tables Va and Vb, or a sequence complementary thereto.

- 5 Other preferred fragments of the invention are polynucleotides comprising polynucleotides encoding domains of polypeptides. Such fragments may be used to obtain other polynucleotides encoding polypeptides having similar domains using hybridization or RT-PCR techniques. Alternatively, these fragments may be used to express a polypeptide domain which may present a specific biological property. Preferred domains for the GENSET polypeptides of the invention are
- 10 described in Table VI. Thus, another object of the invention is an isolated, purified or recombinant polynucleotide encoding a polypeptide consisting of, consisting essentially of, or comprising a contiguous span of at least 5, 6, 8, 10, 12, 15, 20, 25, 30, 35, 40, 50, 60, 75, 100, 150, 200, 250, 300, 350, 400, 450 or 500 consecutive amino acids of a sequence selected from the group consisting of the sequences of SEQ ID Nos: 242-482, to the extent that a contiguous span of these lengths is
- 15 consistent with the lengths of said selected sequence, where said contiguous span comprises at least 1, 2, 3, 5, or 10 of the amino acid positions of a domain of said selected sequence. The present invention also encompasses isolated, purified or recombinant polynucleotides encoding a polypeptide comprising a contiguous span of at least 5, 6, 8, 10, 12, 15, 20, 25, 30, 35, 40, 50, 60, 75, 100, 150, 200, 250, 300, 350, 400, 450 or 500 consecutive amino acids of a sequence selected
- 20 from the group consisting of sequences of SEQ ID Nos: 242-482, to the extent that a contiguous span of these lengths is consistent with the lengths of said selected sequence, where said contiguous span is a domain of said selected sequence. The present invention also encompasses isolated, purified or recombinant polynucleotides encoding a polypeptide comprising a domain of a sequence selected from the group consisting of the sequences of SEQ ID Nos: 242-482.
- 25 The present invention further encompasses any combination of the polynucleotide fragments listed in this section.

Oligonucleotide primers and probes

- The present invention also encompasses fragments of GENSET polynucleotides for use as primers and probes. Polynucleotides derived from the GENSET genomic and cDNA sequences are
- 30 useful in order to detect the presence of at least a copy of a GENSET polynucleotide or fragment, complement, or variant thereof in a test sample.

Structural definition

- Any polynucleotide of the invention may be used as a primer or probe. Particularly preferred probes and primers of the invention include isolated, purified, or recombinant
- 35 polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, 1000, 1500 or 2000 nucleotides of a sequence selected from the group

consisting of the GENSET genomic sequences, the cDNA sequences and the sequences fully complementary thereto. Another object of the invention is a purified, isolated, or recombinant polynucleotide comprising the nucleotide sequence of a sequence selected from the group consisting of the sequences of SEQ ID Nos: 1-241, sequences of clone inserts of the deposited clone pool, sequences fully complementary thereto, allelic variants thereof, and fragments thereof. Moreover, preferred probes and primers of the invention include purified, isolated, or recombinant GENSET cDNAs consisting of, consisting essentially of, or comprising the sequences of SEQ ID Nos: 1-241 and sequences of clone inserts of the deposited clone pool. Particularly preferred probes and primers of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, 1000, 1500 or 2000 nucleotides of a sequence selected from the group consisting of the sequences of SEQ ID Nos: 1-241 and the sequences fully complementary thereto.

Design of primers and probes

A probe or a primer according to the invention has between 8 and 1000 nucleotides in length, or is specified to be at least 12, 15, 18, 20, 25, 35, 40, 50, 60, 70, 80, 100, 250, 500, 1000, 1500 or 2000 nucleotides in length. More particularly, the length of these probes and primers can range from 8, 10, 15, 20, or 30 to 100 nucleotides, preferably from 10 to 50, more preferably from 15 to 30 nucleotides. Shorter probes and primers tend to lack specificity for a target nucleic acid sequence and generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. Longer probes and primers are expensive to produce and can sometimes self-hybridize to form hairpin structures. The appropriate length for primers and probes under a particular set of assay conditions may be empirically determined by one of skill in the art. The formation of stable hybrids depends on the melting temperature (T_m) of the DNA. The T_m depends on the length of the primer or probe, the ionic strength of the solution and the G+C content. The higher the G+C content of the primer or probe, the higher is the melting temperature because G:C pairs are held by three H bonds whereas A:T pairs have only two. The GC content in the probes of the invention usually ranges between 10 and 75 %, preferably between 35 and 60 %, and more preferably between 40 and 55 %.

For amplification purposes, pairs of primers with approximately the same T_m are preferable. Primers may be designed using the OSP software (Hillier and Green, 1991), the disclosure of which is incorporated by reference in its entirety, based on GC content and melting temperatures of oligonucleotides, or using PC-Rare (<http://bioinformatics.weizmann.ac.il/software/PC-Rare/doc/manuel.html>) based on the octamer frequency disparity method (Griffais *et al.*, 1991), the disclosure of which is incorporated by reference in its entirety. DNA amplification techniques are well known to those skilled in the art. Amplification techniques that can be used in the context of the present invention include, but are not limited to, the ligase chain reaction (LCR) described in EP-A- 320 308, WO 9320227 and EP-A-439 182, the

polymerase chain reaction (PCR, RT-PCR) and techniques such as the nucleic acid sequence based amplification (NASBA) described in Guatelli *et al.* (1990) and in Compton (1991), Q-beta amplification as described in European Patent Application No 4544610, strand displacement amplification as described in Walker *et al.* (1996) and EP A 684 315 and, target mediated
5 amplification as described in PCT Publication WO 9322461, the disclosures of which are incorporated by reference in their entireties.

LCR and Gap LCR are exponential amplification techniques, both depend on DNA ligase to join adjacent primers annealed to a DNA molecule. In Ligase Chain Reaction (LCR), probe pairs are used which include two primary (first and second) and two secondary (third and fourth) probes,
10 all of which are employed in molar excess to target. The first probe hybridizes to a first segment of the target strand and the second probe hybridizes to a second segment of the target strand, the first and second segments being contiguous so that the primary probes abut one another in 5' phosphate-3'hydroxyl relationship, and so that a ligase can covalently fuse or ligate the two probes into a fused product. In addition, a third (secondary) probe can hybridize to a portion of the first probe and a
15 fourth (secondary) probe can hybridize to a portion of the second probe in a similar abutting fashion. Of course, if the target is initially double stranded, the secondary probes also will hybridize to the target complement in the first instance. Once the ligated strand of primary probes is separated from the target strand, it will hybridize with the third and fourth probes, which can be ligated to form a complementary, secondary ligated product. It is important to realize that the
20 ligated products are functionally equivalent to either the target or its complement. By repeated cycles of hybridization and ligation, amplification of the target sequence is achieved. A method for multiplex LCR has also been described (WO 9320227), the disclosure of which is incorporated by reference in its entirety. Gap LCR (GLCR) is a version of LCR where the probes are not adjacent but are separated by 2 to 3 bases.

25 For amplification of mRNAs, it is within the scope of the present invention to reverse transcribe mRNA into cDNA followed by polymerase chain reaction (RT-PCR); or, to use a single enzyme for both steps as described in U.S. Patent No. 5,322,770 or, to use Asymmetric Gap LCR (RT-AGLCR) as described by Marshall *et al.* (1994), the disclosures of which are incorporated by reference in its entireties. AGLCR is a modification of GLCR that allows the amplification of
30 RNA.

The PCR technology is the preferred amplification technique used in the present invention. A variety of PCR techniques are familiar to those skilled in the art. For a review of PCR technology, see White (1997), Erlich (1992) and the publication entitled "PCR Methods and Applications" (1991, Cold Spring Harbor Laboratory Press), the disclosures of which are
35 incorporated by reference in its entireties. In each of these PCR procedures, PCR primers on either side of the nucleic acid sequences to be amplified are added to a suitably prepared nucleic acid sample along with dNTPs and a thermostable polymerase such as Taq polymerase, Pfu polymerase,

Tth polymerase or Vent polymerase. The nucleic acid in the sample is denatured and the PCR primers are specifically hybridized to complementary nucleic acid sequences in the sample. The hybridized primers are extended. Thereafter, another cycle of denaturation, hybridization, and extension is initiated. The cycles are repeated multiple times to produce an amplified fragment
5 containing the nucleic acid sequence between the primer sites. PCR has further been described in several patents including US Patents 4,683,195; 4,683,202; and 4,965,188, the disclosures of which are incorporated herein by reference in their entireties.

Preparation of primers and probes

The primers and probes can be prepared by any suitable method, including, for example,
10 cloning and restriction of appropriate sequences and direct chemical synthesis by a method such as the phosphodiester method of Narang *et al.*(1979), the phosphodiester method of Brown *et al.*(1979), the diethylphosphoramidite method of Beaucage *et al.*(1981) and the solid support method described in EP 0 707 592, which disclosures are hereby incorporated by reference in their entireties.

15 Detection probes are generally nucleic acid sequences or uncharged nucleic acid analogs such as, for example peptide nucleic acids which are disclosed in International Patent Application WO 92/20702, morpholino analogs which are described in U.S. Patents Numbered 5,185,444; 5,034,506 and 5,142,047, which disclosures are hereby incorporated by reference in their entireties. The probe may have to be rendered "non-extendable" in that additional dNTPs cannot be added to
20 the probe. In and of themselves analogs usually are non-extendable and nucleic acid probes can be rendered non-extendable by modifying the 3' end of the probe such that the hydroxyl group is no longer capable of participating in elongation. For example, the 3' end of the probe can be functionalized with the capture or detection label to thereby consume or otherwise block the hydroxyl group. Alternatively, the 3' hydroxyl group simply can be cleaved, replaced or modified,
25 U.S. Patent Application Serial No. 07/049,061 filed April 19, 1993, which disclosure is hereby incorporated by reference in its entirety, describes modifications, which can be used to render a probe non-extendable.

Labeling of probes

Any of the polynucleotides of the present invention can be labeled, if desired, by
30 incorporating any label known in the art to be detectable by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include radioactive substances (including, ^{32}P , ^{35}S , ^3H , ^{125}I), fluorescent dyes (including, 5-bromodesoxyuridin, fluorescein, acetylaminofluorene, digoxigenin) or biotin. Preferably, polynucleotides are labeled at their 3' and 5' ends. Examples of non-radioactive labeling of nucleic acid fragments are described
35 in the French patent No. FR-7810975 or by Urdea *et al* (1988) or Sanchez-Pescador *et al* (1988), which disclosures are hereby incorporated by reference in their entireties. In addition, the probes

according to the present invention may have structural characteristics such that they allow the signal amplification, such structural characteristics being, for example, branched DNA probes as those described by Urdea *et al.* in 1991 or in the European patent No. EP 0 225 807 (Chiron), which disclosures are hereby incorporated by reference in their entireties.

5 The detectable probe may be single stranded or double stranded and may be made using techniques known in the art, including *in vitro* transcription, nick translation, or kinase reactions. A nucleic acid sample containing a sequence capable of hybridizing to the labeled probe is contacted with the labeled probe. If the nucleic acid in the sample is double stranded, it may be denatured prior to contacting the probe. In some applications, the nucleic acid sample may be immobilized on
10 a surface such as a nitrocellulose or nylon membrane. The nucleic acid sample may comprise nucleic acids obtained from a variety of sources, including genomic DNA, cDNA libraries, RNA, or tissue samples.

Procedures used to detect the presence of nucleic acids capable of hybridizing to the detectable probe include well known techniques such as Southern blotting, Northern blotting, dot
15 blotting, colony hybridization, and plaque hybridization. In some applications, the nucleic acid capable of hybridizing to the labeled probe may be cloned into vectors such as expression vectors, sequencing vectors, or *in vitro* transcription vectors to facilitate the characterization and expression of the hybridizing nucleic acids in the sample. For example, such techniques may be used to isolate and clone sequences in a genomic library or cDNA library which are capable of hybridizing to the
20 detectable probe as described herein.

Immobilization of probes

A label can also be used to capture the primer, so as to facilitate the immobilization of either the primer or a primer extension product, such as amplified DNA, on a solid support. A capture label is attached to the primers or probes and can be a specific binding member which forms
25 a binding pair with the solid's phase reagent's specific binding member (e.g. biotin and streptavidin). Therefore depending upon the type of label carried by a polynucleotide or a probe, it may be employed to capture or to detect the target DNA. Further, it will be understood that the polynucleotides, primers or probes provided herein, may, themselves, serve as the capture label. For example, in the case where a solid phase reagent's binding member is a nucleic acid sequence,
30 it may be selected such that it binds a complementary portion of a primer or probe to thereby immobilize the primer or probe to the solid phase. In cases where a polynucleotide probe itself serves as the binding member, those skilled in the art will recognize that the probe will contain a sequence or "tail" that is not complementary to the target. In the case where a polynucleotide primer itself serves as the capture label, at least a portion of the primer will be free to hybridize with
35 a nucleic acid on a solid phase. DNA Labeling techniques are well known to the skilled technician.

The probes of the present invention are useful for a number of purposes. They can be notably used in Southern hybridization to genomic DNA. The probes can also be used to detect PCR amplification products. They may also be used to detect mismatches in the GENSET gene or mRNA using other techniques.

5 Any of the polynucleotides, primers and probes of the present invention can be conveniently immobilized on a solid support. The solid support is not critical and can be selected by one skilled in the art. Thus, latex particles, microparticles, magnetic beads, non-magnetic beads (including polystyrene beads), membranes (including nitrocellulose strips), plastic tubes, walls of microtiter wells, glass or silicon chips, sheep (or other suitable animal's) red blood cells and
10 duracytes are all suitable examples. Suitable methods for immobilizing nucleic acids on solid phases include ionic, hydrophobic, covalent interactions and the like. A solid support, as used herein, refers to any material which is insoluble, or can be made insoluble by a subsequent reaction. The solid support can be chosen for its intrinsic ability to attract and immobilize the capture reagent. Alternatively, the solid phase can retain an additional receptor which has the ability to attract and
15 immobilize the capture reagent. The additional receptor can include a charged substance that is oppositely charged with respect to the capture reagent itself or to a charged substance conjugated to the capture reagent. As yet another alternative, the receptor molecule can be any specific binding member which is immobilized upon (attached to) the solid support and which has the ability to immobilize the capture reagent through a specific binding reaction. The receptor molecule enables
20 the indirect binding of the capture reagent to a solid support material before the performance of the assay or during the performance of the assay. The solid phase thus can be a plastic, derivatized plastic, magnetic or non-magnetic metal, glass or silicon surface of a test tube, microtiter well, sheet, bead, microparticle, chip, sheep (or other suitable animal's) red blood cells, duracytes® and other configurations known to those of ordinary skill in the art. The polynucleotides of the
25 invention can be attached to or immobilized on a solid support individually or in groups of at least 2, 5, 8, 10, 12, 15, 20, or 25 distinct polynucleotides of the invention to a single solid support. In addition, polynucleotides other than those of the invention may be attached to the same solid support as one or more polynucleotides of the invention.

Oligonucleotide array

30 A substrate comprising a plurality of oligonucleotide primers or probes of the invention may be used either for detecting or amplifying targeted sequences in GENSET genes, may also be used for detecting mutations in the coding or in the non-coding sequences of GENSET genes, and may also be used to determine GENSET gene expression in different contexts such as in different tissues, at different stages of a process (embryo development, disease treatment), and in patients
35 versus healthy individuals as described elsewhere in the application.

As used herein, the term “array” means a one dimensional, two dimensional, or multidimensional arrangement of nucleic acids of sufficient length to permit specific detection of gene expression. For example, the array may contain a plurality of nucleic acids derived from genes whose expression levels are to be assessed. The array may include a GENSET genomic DNA, a GENSET cDNA, sequences complementary thereto or fragments thereof. Preferably, the fragments are at least 12, 15, 18, 20, 25, 30, 35, 40 or 50 nucleotides in length. More preferably, the fragments are at least 100 nucleotides in length. Even more preferably, the fragments are more than 100 nucleotides in length. In some embodiments the fragments may be more than 500 nucleotides in length.

Any polynucleotide provided herein may be attached in overlapping areas or at random locations on the solid support. Alternatively the polynucleotides of the invention may be attached in an ordered array wherein each polynucleotide is attached to a distinct region of the solid support which does not overlap with the attachment site of any other polynucleotide. Preferably, such an ordered array of polynucleotides is designed to be “addressable” where the distinct locations are recorded and can be accessed as part of an assay procedure. Addressable polynucleotide arrays typically comprise a plurality of different oligonucleotide probes that are coupled to a surface of a substrate in different known locations. The knowledge of the precise location of each polynucleotides location makes these “addressable” arrays particularly useful in hybridization assays. Any addressable array technology known in the art can be employed with the polynucleotides of the invention. One particular embodiment of these polynucleotide arrays is known as the Genechips™, and has been generally described in US Patent 5,143,854; PCT publications WO 90/15070 and 92/10092, which disclosures are hereby incorporated by reference in their entireties. These arrays may generally be produced using mechanical synthesis methods or light directed synthesis methods which incorporate a combination of photolithographic methods and solid phase oligonucleotide synthesis (Fodor *et al.*, 1991), which disclosure is hereby incorporated by reference in its entirety. The immobilization of arrays of oligonucleotides on solid supports has been rendered possible by the development of a technology generally identified as “Very Large Scale Immobilized Polymer Synthesis” (VLSIPS™) in which, typically, probes are immobilized in a high density array on a solid surface of a chip. Examples of VLSIPS™ technologies are provided in US Patents 5,143,854; and 5,412,087 and in PCT Publications WO 90/15070, WO 92/10092 and WO 95/11995, which disclosures are hereby incorporated by reference in their entireties, which describe methods for forming oligonucleotide arrays through techniques such as light-directed synthesis techniques. In designing strategies aimed at providing arrays of nucleotides immobilized on solid supports, further presentation strategies were developed to order and display the oligonucleotide arrays on the chips in an attempt to maximize hybridization patterns and sequence information. Examples of such presentation strategies are disclosed in PCT Publications WO

94/12305, WO 94/11530, WO 97/29212 and WO 97/31256, the disclosures of which are incorporated herein by reference in their entireties.

Consequently, the invention concerns an array of nucleic acid molecules comprising at least one polynucleotide of the invention, particularly a probe or primer as described herein. Preferably, the invention concerns an array of nucleic acid comprising at least two polynucleotides of the invention, particularly probes or primers as described herein. Preferably, the invention concerns an array of nucleic acid comprising at least five polynucleotides of the invention, particularly probes or primers as described herein.

A preferred embodiment of the present invention is an array of polynucleotides of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 100, 500, 1000, 1500 or 2000 nucleotides in length which includes at least 1, 2, 5, 10, 15, 20, 35, 50, 100, 150 or 200 sequences selected from the group consisting of the sequences of SEQ ID Nos: 1-241 and sequences of clone inserts of the deposited clone pool, sequences fully complementary thereto, and fragments thereof.

Methods of making the polynucleotides of the invention

The present invention also comprises methods of making the polynucleotides of the invention, including the polynucleotides of SEQ ID Nos: 1-241, genomic DNA obtainable therefrom, or fragment thereof. These methods comprise sequentially linking together nucleotides to produce the nucleic acids having the preceding sequences. Polynucleotides of the invention may be synthesized either enzymatically using techniques well known to those skilled in the art including amplification or hybridization-based methods as described herein, or chemically.

A variety of chemical methods of synthesizing nucleic acids are known to those skilled in the art. In many of these methods, synthesis is conducted on a solid support. These included the 3' phosphoramidite methods in which the 3' terminal base of the desired oligonucleotide is immobilized on an insoluble carrier. The nucleotide base to be added is blocked at the 5' hydroxyl and activated at the 3' hydroxyl so as to cause coupling with the immobilized nucleotide base. Deblocking of the new immobilized nucleotide compound and repetition of the cycle will produce the desired polynucleotide. Alternatively, polynucleotides may be prepared as described in U.S. Patent No. 5,049,656, which disclosure is hereby incorporated by reference in its entirety. In some embodiments, several polynucleotides prepared as described above are ligated together to generate longer polynucleotides having a desired sequence.

POLYPEPTIDES OF THE INVENTION

The term "GENSET polypeptides" is used herein to embrace all of the proteins and polypeptides of the present invention. The present invention encompasses GENSET polypeptides, including recombinant, isolated or purified GENSET polypeptides consisting of, consisting essentially of, or comprising a sequence selected from the group consisting of SEQ ID Nos: 242-482, the polypeptides encoded by human cDNAs contained in the deposited clones, the mature

proteins included in SEQ ID Nos: 242-272 and 274-384, mature proteins encoded by the clone inserts of the deposited clone pool, and variants thereof. Other objects of the invention are polypeptides encoded by the polynucleotides of the invention as well as fusion polypeptides comprising such polypeptide.

5 Polypeptide variants

The present invention further provides for GENSET polypeptides encoded by allelic and splice variants, orthologs, and/or species homologues. Procedures known in the art can be used to obtain, allelic variants, splice variants, orthologs, and/or species homologues of polynucleotides encoding by polypeptides of the group consisting of SEQ ID Nos: 242-482, mature proteins
10 included in SEQ ID Nos: 242-272 and 274-384, and polypeptides either full-length or mature encoded by the clone inserts of the deposited clone pool, using information from the sequences disclosed herein or the clones deposited with the ATCC.

The polypeptides of the present invention also include polypeptides having an amino acid sequence at least 50% identical, more preferably at least 60% identical, and still more preferably
15 70%, 80%, 90%, 95%, 96%, 97%, 98% or 99% identical to a polypeptide selected from the group consisting of the sequences of SEQ ID Nos: 242-482, mature proteins included in sequences of SEQ ID Nos: 242-272 and 274-384, and full-length or mature polypeptides encoded by the clone inserts of the deposited clone pool. By a polypeptide having an amino acid sequence at least, for example, 95% "identical" to a query amino acid sequence of the present invention, it is intended that the
20 amino acid sequence of the subject polypeptide is identical to the query sequence except that the subject polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the query amino acid sequence. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a query amino acid sequence, up to 5% (5 of 100) of the amino acid residues in the subject sequence may be inserted, deleted, (indels) or substituted with
25 another amino acid.

Further polypeptides of the present invention include polypeptides which have at least 90% similarity, more preferably at least 95% similarity, and still more preferably at least 96%, 97%, 98% or 99% similarity to those described above. By a polypeptide having an amino acid sequence at least, for example, 95% "similar" to a query amino acid sequence of the present invention, it is
30 intended that the amino acid sequence of the subject polypeptide is similar (i.e. contain identical or equivalent amino acid residues) to the query sequence except that the subject polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the query amino acid sequence. In other words, to obtain a polypeptide having an amino acid sequence at least 95% similar to a query amino acid sequence, up to 5% (5 of 100) of the amino acid residues in the
35 subject sequence may be inserted, deleted, (indels) or substituted with another non-equivalent amino acid.

These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence. The query sequence may be an entire amino acid sequence selected from the group consisting of sequences of SEQ ID Nos: 242-482 and those encoded by the clone inserts of the deposited clone pool or any fragment specified as described herein.

The variant polypeptides described herein are included in the present invention regardless of whether they have their normal biological activity. This is because even where a particular polypeptide molecule does not have biological activity, one of skill in the art would still know how to use the polypeptide, for instance, as a vaccine or to generate antibodies. Other uses of the polypeptides of the present invention that do not have GENSET biological activity include, *inter alia*, as epitope tags, in epitope mapping, and as molecular weight markers on SDS-PAGE gels or on molecular sieve gel filtration columns using methods known to those of skill in the art. As described below, the polypeptides of the present invention can also be used to raise polyclonal and monoclonal antibodies, which are useful in assays for detecting GENSET protein expression or as agonists and antagonists capable of enhancing or inhibiting GENSET protein function. Further, such polypeptides can be used in the yeast two-hybrid system to "capture" GENSET protein binding proteins, which are also candidate agonists and antagonists according to the present invention (*See, e.g., Fields et al. 1989*), which disclosure is hereby incorporated by reference in its entirety.

Preparation of the polypeptides of the invention

The polypeptides of the present invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. The polypeptides of the present invention are preferably provided in an isolated form, and may be partially or preferably substantially purified.

Consequently, the present invention also comprises methods of making the polypeptides of the invention, particularly polypeptides encoded by the cDNAs of SEQ ID Nos: 1-241, mature proteins encoded by fragments of SEQS ID Nos: 1-31 and 33-143, full-length and mature polypeptides encoded by the clone inserts of the deposited clone pool, genomic DNA obtainable therefrom, or fragments thereof and methods of making the polypeptides of SEQ ID Nos: 242-482, mature polypeptides included in SEQ ID Nos: 242-272 and 274-384, or fragments thereof. The methods comprise sequentially linking together amino acids to produce the nucleic polypeptides having the preceding sequences. In some embodiments, the polypeptides made by these methods are 150 amino acids or less in length. In other embodiments, the polypeptides made by these methods are 120 amino acids or less in length.

Isolation*From natural sources*

The GENSET proteins of the invention may be isolated from natural sources, including bodily fluids, tissues and cells, whether directly isolated or cultured cells, of humans or non-human animals. Methods for extracting and purifying natural proteins are known in the art, and include the use of detergents or chaotropic agents to disrupt particles followed by differential extraction and separation of the polypeptides by ion exchange chromatography, affinity chromatography, sedimentation according to density, and gel electrophoresis. See, for example, "Methods in Enzymology, Academic Press, 1993" for a variety of methods for purifying proteins, which disclosure is hereby incorporated by reference in its entirety. Polypeptides of the invention also can be purified from natural sources using antibodies directed against the polypeptides of the invention, such as those described herein, in methods which are well known in the art of protein purification.

From recombinant sources

Preferably, the GENSET polypeptides of the invention are recombinantly produced using routine expression methods known in the art. The polynucleotide encoding the desired polypeptide is operably linked to a promoter into an expression vector suitable for any convenient host. Both eukaryotic and prokaryotic host systems are used in forming recombinant polypeptides. The polypeptide is then isolated from lysed cells or from the culture medium and purified to the extent needed for its intended use.

Any GENSET polynucleotide, including those described in SEQ ID Nos: 1-241, those of clone inserts of the deposited clone pool, and allelic variants thereof may be used to express GENSET polypeptides. The nucleic acid encoding the GENSET polypeptide to be expressed is operably linked to a promoter in an expression vector using conventional cloning technology. The GENSET insert in the expression vector may comprise the full coding sequence for the GENSET protein or a portion thereof, especially the sequence for a mature polypeptide. For example, the GENSET derived insert may encode a polypeptide comprising at least 6, 8, 10, 12, 15, 20, 25, 30, 35, 40, 50, 60, 75, 100, 150 or 200 consecutive amino acids of a GENSET protein selected from the group consisting of sequences of SEQ ID Nos: 242-482 and polypeptides encoded by the clone inserts of the deposited clone pool.

Consequently, a further embodiment of the present invention is a method of making a polypeptide comprising a protein selected from the group consisting of sequences of SEQ ID Nos: 242-482 and polypeptides encoded by the clone inserts of the deposited clone pool, said method comprising the steps of

a) obtaining a cDNA comprising a sequence selected from the group consisting of i) the sequences SEQ ID Nos: 1-241, ii) the sequences of clone inserts of the deposited clone pool one, iii) sequences encoding one of the polypeptide of SEQ ID Nos: 242-482, and iv) sequences of

polynucleotides encoding a polypeptide which is encoded by one of the clone insert of the deposited clone pool;

b) inserting said cDNA in an expression vector such that the cDNA is operably linked to a promoter; and

5 c) introducing said expression vector into a host cell whereby said host cell produces said polypeptide.

In one aspect of this embodiment, the method further comprises the step of isolating the polypeptide. Another embodiment of the present invention is a polypeptide obtainable by the method described in the preceding paragraph.

10 The expression vector is any of the mammalian, yeast, insect or bacterial expression systems known in the art. Commercially available vectors and expression systems are available from a variety of suppliers including Genetics Institute (Cambridge, MA), Stratagene (La Jolla, California), Promega (Madison, Wisconsin), and Invitrogen (San Diego, California). If desired, to enhance expression and facilitate proper protein folding, the codon context and codon pairing of the
15 sequence is optimized for the particular expression organism in which the expression vector is introduced, as explained in U.S. Patent No. 5,082,767, which disclosure is hereby incorporated by reference in its entirety.

In one embodiment, the entire coding sequence of a GENSET cDNA and the 3'UTR through the poly A signal of the cDNA is operably linked to a promoter in the expression vector.
20 Alternatively, if the nucleic acid encoding a portion of the GENSET protein lacks a methionine to serve as the initiation site, an initiating methionine can be introduced next to the first codon of the nucleic acid using conventional techniques. Similarly, if the insert from the GENSET cDNA lacks a poly A signal, this sequence can be added to the construct by, for example, splicing out the Poly A signal from pSG5 (Stratagene) using BglII and SalI restriction endonuclease enzymes and
25 incorporating it into the mammalian expression vector pXT1 (Stratagene). pXT1 contains the LTRs and a portion of the gag gene from Moloney Murine Leukemia Virus. The position of the LTRs in the construct allow efficient stable transfection. The vector includes the Herpes Simplex Thymidine Kinase promoter and the selectable neomycin gene. The nucleic acid encoding the GENSET protein or a portion thereof is obtained by PCR from a vector containing a GENSET cDNA selected
30 from the group consisting of the sequences of SEQ ID Nos: 1-241 and the clone inserts of the deposited clone pool using oligonucleotide primers complementary to the GENSET cDNA or portion thereof and containing restriction endonuclease sequences for Pst I incorporated into the 5' primer and BglII at the 5' end of the corresponding cDNA 3' primer, taking care to ensure that the sequence encoding the GENSET protein or a portion thereof is positioned properly with respect to
35 the poly A signal. The purified fragment obtained from the resulting PCR reaction is digested with PstI, blunt ended with an exonuclease, digested with Bgl II, purified and ligated to pXT1, now containing a poly A signal and digested with BglII.

Alternatively, cDNAs encoding secreted proteins may be cloned into pED6dpc2 (DiscoverEase, Genetics Institute, Cambridge, MA). The resulting pED6dpc2 constructs may be transfected into a suitable host cell, such as COS 1 cells. Methotrexate resistant cells are selected and expanded. Preferably, the secreted protein expressed from the cDNA is released into the culture medium thereby facilitating purification.

In another embodiment, it is often advantageous to add to the recombinant polynucleotide additional nucleotide sequence which codes for secretory or leader sequences, pro-sequences, sequences which aid in purification, such as multiple histidine residues, or an additional sequence for stability during recombinant production.

As a control, the expression vector lacking a cDNA insert is introduced into host cells or organisms.

Transfection of a GENSET expressing vector into mouse NTH 3T3 cells is but one embodiment of introducing polynucleotides into host cells. Introduction of a polynucleotide encoding a polypeptide into a host cell can be effected by calcium phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, or other methods. Such methods are described in many standard laboratory manuals, such as Davis *et al.* (1986), which disclosure is hereby incorporated by reference in its entirety. It is specifically contemplated that the polypeptides of the present invention may in fact be expressed by a host cell lacking a recombinant vector.

Recombinant cell extracts, or proteins from the culture medium if the expressed polypeptide is secreted, are then prepared and proteins separated by gel electrophoresis. If desired, the proteins may be ammonium sulfate precipitated or separated based on size or charge prior to electrophoresis. The proteins present are detected using techniques such as Coomassie or silver staining or using antibodies against the protein encoded by the GENSET cDNA of interest. Coomassie and silver staining techniques are familiar to those skilled in the art.

Proteins from the host cells or organisms containing an expression vector which contains the GENSET cDNA or a fragment thereof are compared to those from the control cells or organism. The presence of a band from the cells containing the expression vector which is absent in control cells indicates that the GENSET cDNA is expressed. Generally, the band corresponding to the protein encoded by the GENSET cDNA will have a mobility near that expected based on the number of amino acids in the open reading frame of the cDNA. However, the band may have a mobility different than that expected as a result of modifications such as glycosylation, ubiquitination, or enzymatic cleavage.

Alternatively, the GENSET polypeptide to be expressed may also be a product of transgenic animals, i.e., as a component of the milk of transgenic cows, goats, pigs or sheeps which are characterized by somatic or germ cells containing a nucleotide sequence encoding the protein of interest.

A polypeptide of this invention can be recovered and purified from recombinant cell cultures by well-known methods including differential extraction, ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. See, for example, "Methods in Enzymology", *supra* for a variety of methods for purifying proteins. Most preferably, high performance liquid chromatography ("HPLC") is employed for purification. A recombinantly produced version of a GENSET polypeptide can be substantially purified using techniques described herein or otherwise known in the art, such as, for example, by the one-step method described in Smith and Johnson (1988), which disclosure is hereby incorporated by reference in its entirety. Polypeptides of the invention also can be purified from recombinant sources using antibodies directed against the polypeptides of the invention, such as those described herein, in methods which are well known in the art of protein purification.

Preferably, the recombinantly expressed GENSET polypeptide is purified using standard immunochromatography techniques such as the one described in the section entitled "Immunoaffinity Chromatography". In such procedures, a solution containing the protein of interest, such as the culture medium or a cell extract, is applied to a column having antibodies against the protein attached to the chromatography matrix. The recombinant protein is allowed to bind the immunochromatography column. Thereafter, the column is washed to remove non-specifically bound proteins. The specifically bound protein is then released from the column and recovered using standard techniques.

If antibody production is not possible, the GENSET cDNA sequence or fragment thereof may be incorporated into expression vectors designed for use in purification schemes employing chimeric polypeptides. In such strategies the coding sequence of the GENSET cDNA or fragment thereof is inserted in frame with the gene encoding the other half of the chimera. The other half of the chimera may be beta-globin or a nickel binding polypeptide encoding sequence. A chromatography matrix having antibody to beta-globin or nickel attached thereto is then used to purify the chimeric protein. Protease cleavage sites may be engineered between the beta-globin gene or the nickel binding polypeptide and the GENSET cDNA or fragment thereof. Thus, the two polypeptides of the chimera may be separated from one another by protease digestion.

One useful expression vector for generating beta-globin chimerics is pSG5 (Stratagene), which encodes rabbit beta-globin. Intron II of the rabbit beta-globin gene facilitates splicing of the expressed transcript, and the polyadenylation signal incorporated into the construct increases the level of expression. These techniques as described are well known to those skilled in the art of molecular biology. Standard methods are published in methods texts such as Davis *et al.*, (1986) and many of the methods are available from Stratagene, Life Technologies, Inc., or Promega.

Polypeptide may additionally be produced from the construct using *in vitro* translation systems such as the *In vitro* ExpressTM Translation Kit (Stratagene).

Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. In addition, 5 polypeptides of the invention may also include an initial modified methionine residue, in some cases as a result of host-mediated processes. Thus, it is well known in the art that the N-terminal methionine encoded by the translation initiation codon generally is removed with high efficiency from any protein after translation in all eukaryotic cells. While the N-terminal methionine on most proteins also is efficiently removed in most prokaryotes, for some proteins, this prokaryotic removal 10 process is inefficient, depending on the nature of the amino acid to which the N-terminal methionine is covalently linked.

From chemical synthesis

In addition, polypeptides of the invention, especially short protein fragments, can be chemically synthesized using techniques known in the art (*See, e.g.*, Creighton, 1983; and 15 Hunkapiller *et al.*, 1984), which disclosures are hereby incorporated by reference in their entireties. For example, a polypeptide corresponding to a fragment of a polypeptide sequence of the invention can be synthesized by use of a peptide synthesizer. A variety of methods of making polypeptides are known to those skilled in the art, including methods in which the carboxyl terminal amino acid is bound to polyvinyl benzene or another suitable resin. The amino acid to be added possesses 20 blocking groups on its amino moiety and any side chain reactive groups so that only its carboxyl moiety can react. The carboxyl group is activated with carbodiimide or another activating agent and allowed to couple to the immobilized amino acid. After removal of the blocking group, the cycle is repeated to generate a polypeptide having the desired sequence. Alternatively, the methods described in U.S. Patent No. 5,049,656, which disclosure is hereby incorporated by reference in its 25 entirety, may be used.

Furthermore, if desired, nonclassical amino acids or chemical amino acid analogs can be introduced as a substitution or addition into the polypeptide sequence. Non-classical amino acids include, but are not limited to, to the D-isomers of the common amino acids, 2,4-diaminobutyric acid, α -amino isobutyric acid, 4-aminobutyric acid, Abu, 2-amino butyric acid, γ -Abu, ϵ -Ahx, 6- 30 amino hexanoic acid, Aib, 2-amino isobutyric acid, 3-amino propionic acid, ornithine, norleucine, norvaline, hydroxyproline, sarcosine, citrulline, homocitrulline, cysteic acid, t-butylglycine, t-butylalanine, phenylglycine, cyclohexylalanine, β -alanine, fluoroamino acids, designer amino acids such as β -methyl amino acids, Ca-methyl amino acids, Na-methyl amino acids, and amino acid analogs in general. Furthermore, the amino acid can be D (dextrorotary) or L (levorotary).

Modifications

The invention encompasses polypeptides which are differentially modified during or after translation, e.g., by glycosylation, acetylation, phosphorylation, amidation, derivatization by known protecting/blocking groups, proteolytic cleavage, linkage to an antibody molecule or other cellular ligand, etc. Any of numerous chemical modifications may be carried out by known techniques, including but not limited, to specific chemical cleavage by cyanogen bromide, trypsin, chymotrypsin, papain, V8 protease, NaBH₄; acetylation, formylation, oxidation, reduction; metabolic synthesis in the presence of tunicamycin; etc.

Additional post-translational modifications encompassed by the invention include, for example, e.g., N-linked or O-linked carbohydrate chains, processing of N-terminal or C-terminal ends), attachment of chemical moieties to the amino acid backbone, chemical modifications of N-linked or O-linked carbohydrate chains, and addition or deletion of an N-terminal methionine residue as a result of prokaryotic host cell expression. The polypeptides may also be modified with a detectable label, such as an enzymatic, fluorescent, isotopic or affinity label to allow for detection and isolation of the protein.

Also provided by the invention are chemically modified derivatives of the polypeptides of the invention which may provide additional advantages such as increased solubility, stability and circulating time of the polypeptide, or decreased immunogenicity. See U.S. Patent No: 4,179,337. The chemical moieties for derivatization may be selected See U.S. Patent NO: 4,179,337, which disclosure is hereby incorporated by reference in its entirety. The chemical moieties for derivatization may be selected from water soluble polymers such as polyethylene glycol, ethylene glycol/propylene glycol copolymers, carboxymethylcellulose, dextran, polyvinyl alcohol and the like. The polypeptides may be modified at random positions within the molecule, or at predetermined positions within the molecule and may include one, two, three or more attached chemical moieties.

The polymer may be of any molecular weight, and may be branched or unbranched. For polyethylene glycol, the preferred molecular weight is between about 1 kDa and about 100 kDa (the term "about" indicating that in preparations of polyethylene glycol, some molecules will weigh more, some less, than the stated molecular weight) for ease in handling and manufacturing. Other sizes may be used, depending on the desired therapeutic profile (e.g., the duration of sustained release desired, the effects, if any on biological activity, the ease in handling, the degree or lack of antigenicity and other known effects of the polyethylene glycol to a therapeutic protein or analog).

The polyethylene glycol molecules (or other chemical moieties) should be attached to the protein with consideration of effects on functional or antigenic domains of the protein. There are a number of attachment methods available to those skilled in the art, e.g., EP 0 401 384, (coupling PEG to G-CSF), and Malik *et al.* (1992) (reporting pegylation of GM-CSF using tresyl chloride), which disclosures are hereby incorporated by reference in their entireties. For example,

polyethylene glycol may be covalently bound through amino acid residues via a reactive group, such as, a free amino or carboxyl group. Reactive groups are those to which an activated polyethylene glycol molecule may be bound. The amino acid residues having a free amino group may include lysine residues and the N-terminal amino acid residues; those having a free carboxyl group may include aspartic acid residues glutamic acid residues and the C-terminal amino acid residue. Sulfhydryl groups may also be used as a reactive group for attaching the polyethylene glycol molecules. Preferred for therapeutic purposes is attachment at an amino group, such as attachment at the N-terminus or lysine group.

One may specifically desire proteins chemically modified at the N-terminus. Using polyethylene glycol as an illustration of the present composition, one may select from a variety of polyethylene glycol molecules (by molecular weight, branching, etc.), the proportion of polyethylene glycol molecules to protein (polypeptide) molecules in the reaction mix, the type of pegylation reaction to be performed, and the method of obtaining the selected N-terminally pegylated protein. The method of obtaining the N-terminally pegylated preparation (i.e., separating this moiety from other monopegylated moieties if necessary) may be by purification of the N-terminally pegylated material from a population of pegylated protein molecules. Selective proteins chemically modified at the N-terminus modification may be accomplished by reductive alkylation, which exploits differential reactivity of different types of primary amino groups (lysine versus the N-terminal) available for derivatization in a particular protein. Under the appropriate reaction conditions, substantially selective derivatization of the protein at the N-terminus with a carbonyl group containing polymer is achieved.

Multimerization

The polypeptides of the invention may be in monomers or multimers (i.e., dimers, trimers, tetramers and higher multimers). Accordingly, the present invention relates to monomers and multimers of the polypeptides of the invention, their preparation, and compositions containing them. In specific embodiments, the polypeptides of the invention are monomers, dimers, trimers or tetramers. In additional embodiments, the multimers of the invention are at least dimers, at least trimers, or at least tetramers.

Multimers encompassed by the invention may be homomers or heteromers. As used herein, the term “homomer”, refers to a multimer containing only polypeptides corresponding to the amino acid sequences of SEQ ID Nos: 242-482 or encoded by the clone inserts of the deposited clone pool (including fragments, variants, splice variants, and fusion proteins, corresponding to these polypeptides as described herein). These homomers may contain polypeptides having identical or different amino acid sequences. In a specific embodiment, a homomer of the invention is a multimer containing only polypeptides having an identical amino acid sequence. In another specific embodiment, a homomer of the invention is a multimer containing polypeptides having different

amino acid sequences. In specific embodiments, the multimer of the invention is a homodimer (*e.g.*, containing polypeptides having identical or different amino acid sequences) or a homotrimer (*e.g.*, containing polypeptides having identical and/or different amino acid sequences). In additional embodiments, the homomeric multimer of the invention is at least a homodimer, at least a

5 homotrimer, or at least a homotetramer.

As used herein, the term “heteromer” refers to a multimer containing one or more heterologous polypeptides (*i.e.*, polypeptides of different proteins) in addition to the polypeptides of the invention. In a specific embodiment, the multimer of the invention is a heterodimer, a heterotrimer, or a heterotetramer. In additional embodiments, the heteromeric multimer of the

10 invention is at least a heterodimer, at least a heterotrimer, or at least a heterotetramer.

Multimers of the invention may be the result of hydrophobic, hydrophilic, ionic and/or covalent associations and/or may be indirectly linked, by for example, liposome formation. Thus, in one embodiment, multimers of the invention, such as, for example, homodimers or homotrimers, are formed when polypeptides of the invention contact one another in solution. In another

15 embodiment, heteromultimers of the invention, such as, for example, heterotrimers or heterotetramers, are formed when polypeptides of the invention contact antibodies to the polypeptides of the invention (including antibodies to the heterologous polypeptide sequence in a fusion protein of the invention) in solution. In other embodiments, multimers of the invention are formed by covalent associations with and/or between the polypeptides of the invention. Such

20 covalent associations may involve one or more amino acid residues contained in the polypeptide sequence (*e.g.*, that recited in the sequence listing, or contained in the polypeptide encoded by a deposited clone). In one instance, the covalent associations are cross-linking between cysteine residues located within the polypeptide sequences, which interact in the native (*i.e.*, naturally occurring) polypeptide. In another instance, the covalent associations are the consequence of

25 chemical or recombinant manipulation. Alternatively, such covalent associations may involve one or more amino acid residues contained in the heterologous polypeptide sequence in a fusion protein of the invention.

In one example, covalent associations are between the heterologous sequence contained in a fusion protein of the invention (see, *e.g.*, US Patent Number 5,478,925, which disclosure is hereby

30 incorporated by reference in its entirety). In a specific example, the covalent associations are between the heterologous sequence contained in an Fc fusion protein of the invention (as described herein). In another specific example, covalent associations of fusion proteins of the invention are between heterologous polypeptide sequence from another protein that is capable of forming covalently associated multimers, such as for example, osteoprotegerin (see, *e.g.*, International

35 Publication No: WO 98/49305, the contents of which are herein incorporated by reference in its entirety). In another embodiment, two or more polypeptides of the invention are joined through peptide linkers. Examples include those peptide linkers described in U.S. Pat. No. 5,073,627

(hereby incorporated by reference). Proteins comprising multiple polypeptides of the invention separated by peptide linkers may be produced using conventional recombinant DNA technology.

Another method for preparing multimer polypeptides of the invention involves use of polypeptides of the invention fused to a leucine zipper or isoleucine zipper polypeptide sequence.

- 5 Leucine zipper and isoleucine zipper domains are polypeptides that promote multimerization of the proteins in which they are found. Leucine zippers were originally identified in several DNA-binding proteins, and have since been found in a variety of different proteins (Landschulz *et al.*, 1988). Among the known leucine zippers are naturally occurring peptides and derivatives thereof that dimerize or trimerize. Examples of leucine zipper domains suitable for producing
- 10 soluble multimeric proteins of the invention are those described in PCT application WO 94/10308, hereby incorporated by reference. Recombinant fusion proteins comprising a polypeptide of the invention fused to a polypeptide sequence that dimerizes or trimerizes in solution are expressed in suitable host cells, and the resulting soluble multimeric fusion protein is recovered from the culture supernatant using techniques known in the art.
- 15 Trimeric polypeptides of the invention may offer the advantage of enhanced biological activity. Preferred leucine zipper moieties and isoleucine moieties are those that preferentially form trimers. One example is a leucine zipper derived from lung surfactant protein D (SPD), as described in Hoppe *et al.* (1994) and in U.S. patent application Ser. No. 08/446,922, which disclosure is hereby incorporated by reference in its entirety. Other peptides derived from naturally
- 20 occurring trimeric proteins may be employed in preparing trimeric polypeptides of the invention. In another example, proteins of the invention are associated by interactions between Flag® polypeptide sequence contained in fusion proteins of the invention containing Flag® polypeptide sequence. In a further embodiment, associations proteins of the invention are associated by interactions between heterologous polypeptide sequence contained in Flag® fusion proteins of the
- 25 invention and anti Flag® antibody.

- The multimers of the invention may be generated using chemical techniques known in the art. For example, polypeptides desired to be contained in the multimers of the invention may be chemically cross-linked using linker molecules and linker molecule length optimization techniques known in the art (see, e.g., US Patent Number 5,478,925, which is herein incorporated by
- 30 reference in its entirety). Additionally, multimers of the invention may be generated using techniques known in the art to form one or more inter-molecule cross-links between the cysteine residues located within the sequence of the polypeptides desired to be contained in the multimer (see, e.g., US Patent Number 5,478,925, which is herein incorporated by reference in its entirety). Further, polypeptides of the invention may be routinely modified by the addition of cysteine or
- 35 biotin to the C terminus or N-terminus of the polypeptide and techniques known in the art may be applied to generate multimers containing one or more of these modified polypeptides (see, e.g., US Patent Number 5,478,925, which is herein incorporated by reference in its entirety). Additionally,

30 techniques known in the art may be applied to generate liposomes containing the polypeptide components desired to be contained in the multimer of the invention (see, e.g., US Patent Number 5,478,925, which is herein incorporated by reference in its entirety).

Alternatively, multimers of the invention may be generated using genetic engineering techniques known in the art. In one embodiment, polypeptides contained in multimers of the invention are produced recombinantly using fusion protein technology described herein or otherwise known in the art (see, e.g., US Patent Number 5,478,925, which is herein incorporated by reference in its entirety). In a specific embodiment, polynucleotides coding for a homodimer of the invention are generated by ligating a polynucleotide sequence encoding a polypeptide of the invention to a sequence encoding a linker polypeptide and then further to a synthetic polynucleotide encoding the translated product of the polypeptide in the reverse orientation from the original C-terminus to the N-terminus (lacking the leader sequence) (see, e.g., US Patent Number 5,478,925, which is herein incorporated by reference in its entirety). In another embodiment, recombinant techniques described herein or otherwise known in the art are applied to generate recombinant polypeptides of the invention which contain a transmembrane domain (or hydrophobic or signal peptide) and which can be incorporated by membrane reconstitution techniques into liposomes (see, e.g., US Patent Number 5,478,925, which is herein incorporated by reference in its entirety).

Mutated polypeptides

To improve or alter the characteristics of GENSET polypeptides of the present invention, protein engineering may be employed. Recombinant DNA technology known to those skilled in the art can be used to create novel mutant proteins or muteins including single or multiple amino acid substitutions, deletions, additions, or fusion proteins. Such modified polypeptides can show, e.g., increased/decreased biological activity or increased/decreased stability. In addition, they may be purified in higher yields and show better solubility than the corresponding natural polypeptide, at least under certain purification and storage conditions. Further, the polypeptides of the present invention may be produced as multimers including dimers, trimers and tetramers. Multimerization may be facilitated by linkers or recombinantly through heterologous polypeptides such as Fc regions.

N- and C-terminal deletions

It is known in the art that one or more amino acids may be deleted from the N-terminus or C-terminus without substantial loss of biological function. For instance, Ron *et al.* (1993), reported modified KGF proteins that had heparin binding activity even if 3, 8, or 27 N-terminal amino acid residues were missing. Accordingly, the present invention provides polypeptides having one or more residues deleted from the amino terminus of the polypeptides of SEQ ID Nos: 242-482 or that encoded by the clone inserts of the deposited clone pool. Similarly, many examples of biologically functional C-terminal deletion mutants are known. For instance, Interferon gamma shows up to ten times higher activities by deleting 810 amino acid residues from the C-terminus of the protein (See,

e.g., Dobeli, *et al.* 1988), which disclosure is hereby incorporated by reference in its entirety. Accordingly, the present invention provides polypeptides having one or more residues deleted from the carboxy terminus of the polypeptides shown of SEQ ID Nos: 242-482 or encoded by the clone inserts of the deposited clone pool. The invention also provides polypeptides having one or more amino acids deleted from both the amino and the carboxyl termini as described below.

Other mutations

Other mutants in addition to N- and C-terminal deletion forms of the protein discussed above are included in the present invention. It also will be recognized by one of ordinary skill in the art that some amino acid sequences of the GENSET polypeptides of the present invention can be varied without significant effect of the structure or function of the protein. If such differences in sequence are contemplated, it should be remembered that there will be critical areas on the protein which determine activity. Thus, the invention further includes variations of the GENSET polypeptides which show substantial GENSET polypeptide activity. Such mutants include deletions, insertions, inversions, repeats, and substitutions selected according to general rules known in the art so as to have little effect on activity. For example, guidance concerning how to make phenotypically silent amino acid substitutions is provided.

There are two main approaches for studying the tolerance of an amino acid sequence to change (See, Bowie *et al.* 1994), which disclosure is hereby incorporated by reference in its entirety. The first method relies on the process of evolution, in which mutations are either accepted or rejected by natural selection.

The second approach uses genetic engineering to introduce amino acid changes at specific positions of a cloned gene and selections or screens to identify sequences that maintain functionality. These studies have revealed that proteins are surprisingly tolerant of amino acid substitutions. The studies indicate which amino acid changes are likely to be permissive at a certain position of the protein. For example, most buried amino acid residues require nonpolar side chains, whereas few features of surface side chains are generally conserved. Other such phenotypically silent substitutions are described by Bowie *et al.* (*supra*) and the references cited therein.

Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu and Phe; interchange of the hydroxyl residues Ser and Thr, exchange of the acidic residues Asp and Glu, substitution between the amide residues Asn and Gln, exchange of the basic residues Lys and Arg and replacements among the aromatic residues Phe, Tyr. Thus, the fragment, derivative, analog, or homologue of the polypeptide of the present invention may be, for example: (i) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code; or (ii) one in which one or more of the amino acid residues includes a substituent group; or

(iii) one in which the GENSET polypeptide is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol): or (iv) one in which the additional amino acids are fused to the above form of the polypeptide, such as an IgG Fc fusion region peptide or leader or secretory sequence or a sequence which is employed for purification of the above form of the polypeptide or a pro-protein sequence. Such fragments, derivatives and analogs are deemed to be within the scope of those skilled in the art from the teachings herein.

Thus, the GENSET polypeptides of the present invention may include one or more amino acid substitutions, deletions, or additions, either from natural mutations or human manipulation. As indicated, changes are preferably of a minor nature, such as conservative amino acid substitutions that do not significantly affect the folding or activity of the protein. The following groups of amino acids generally represent equivalent changes: (1) Ala, Pro, Gly, Glu, Asp, Gln, Asn, Ser, Thr; (2) Cys, Ser, Tyr, Thr; (3) Val, Ile, Leu, Met, Ala, Phe; (4) Lys, Arg, His; (5) Phe, Tyr, Trp, His.

A specific embodiment of a modified GENSET peptide molecule of interest according to the present invention, includes, but is not limited to, a peptide molecule which is resistant to proteolysis, is a peptide in which the -CONH- peptide bond is modified and replaced by a (CH₂NH) reduced bond, a (NHCO) retro inverso bond, a (CH₂-O) methylene-oxy bond, a (CH₂-S) thiomethylene bond, a (CH₂CH₂) carba bond, a (CO-CH₂) cetomethylene bond, a (CHOH-CH₂) hydroxyethylene bond, a (N-N) bound, a E-alcene bond or also a -CH=CH- bond. The invention also encompasses a human GENSET polypeptide or a fragment or a variant thereof in which at least one peptide bond has been modified as described above.

Amino acids in the GENSET proteins of the present invention that are essential for function can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (*See, e.g.,* Cunningham *et al.* 1989), which disclosure is hereby incorporated by reference in its entirety. The latter procedure introduces single alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for biological activity using assays appropriate for measuring the function of the particular protein. Of special interest are substitutions of charged amino acids with other charged or neutral amino acids which may produce proteins with highly desirable improved characteristics, such as less aggregation. Aggregation may not only reduce activity but also be problematic when preparing pharmaceutical formulations, because aggregates can be immunogenic, (*See, e.g.,* Pinckard *et al.*, 1967; Robbins, *et al.*, 1987; and Cleland, *et al.*, 1993).

A further embodiment of the invention relates to a polypeptide which comprises the amino acid sequence of a GENSET polypeptide having an amino acid sequence which contains at least one conservative amino acid substitution, but not more than 50 conservative amino acid substitutions, not more than 40 conservative amino acid substitutions, not more than 30 conservative amino acid substitutions, and not more than 20 conservative amino acid substitutions. Also provided are

polypeptides which comprise the amino acid sequence of a GENSET polypeptide, having at least one, but not more than 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 conservative amino acid substitutions.

Polypeptide fragments

Structural definition

5 The present invention is further directed to fragments of the amino acid sequences described herein such as the polypeptides of SEQ ID Nos: 242-482, mature polypeptides included in SEQ ID Nos: 242-272 and 274-384, or full-length or mature polypeptides encoded by the clone inserts of the deposited clone pool. More specifically, the present invention embodies purified, isolated, and recombinant polypeptides comprising at least 6, preferably at least 8 to 10, more preferably 12, 15,
10 20, 25, 30, 35, 40, 50, 60, 75, 100, 125, 150, 175, 200, 225, 250, 275, 300, 350, 400, 450 or 500 consecutive amino acids of a polypeptide selected from the group consisting of the sequences of SEQ ID Nos: 242-482, mature polypeptides included in SEQ ID Nos: 242-272 and 274-384, and full-length or mature polypeptides encoded by the clone inserts of the deposited clone pool, and other polypeptides of the present invention.

15 In addition to the above polypeptide fragments, further preferred sub-genuses of polypeptides comprise at least 6 amino acids, wherein "at least 6" is defined as any integer between 6 and the integer representing the C-terminal amino acid of the polypeptide of the present invention including the polypeptide sequences of the sequence listing below. Further included are species of polypeptide fragments at least 6 amino acids in length, as described above, that are further specified
20 in terms of their N-terminal and C-terminal positions. However, included in the present invention as individual species are all polypeptide fragments, at least 6 amino acids in length, as described above, and may be particularly specified by a N-terminal and C-terminal position. That is, every combination of a N-terminal and C-terminal position that a fragment at least 6 contiguous amino acid residues in length could occupy, on any given amino acid sequence of the sequence listing or
25 of the present invention is included in the present invention

 The present invention also provides for the exclusion of any fragment species specified by N-terminal and C-terminal positions or of any fragment sub-genus specified by size in amino acid residues as described above. Any number of fragments specified by N-terminal and C-terminal positions or by size in amino acid residues as described above may be excluded as individual
30 species.

 The above polypeptide fragments of the present invention can be immediately envisaged using the above description and are therefore not individually listed solely for the purpose of not unnecessarily lengthening the specification. Moreover, the above fragments need not have a GENSET biological activity, although polypeptides having these activities are preferred
35 embodiments of the invention, since they would be useful, for example, in immunoassays, in epitope mapping, epitope tagging, as vaccines, and as molecular weight markers. The above

fragments may also be used to generate antibodies to a particular portion of the polypeptide. These antibodies can then be used in immunoassays well known in the art to distinguish between human and non-human cells and tissues or to determine whether cells or tissues in a biological sample are or are not of the same type which express the polypeptides of the present invention.

- 5 It is noted that the above species of polypeptide fragments of the present invention may alternatively be described by the formula "a to b"; where "a" equals the N-terminal most amino acid position and "b" equals the C-terminal most amino acid position of the polynucleotide; and further where "a" equals an integer between 1 and the number of amino acids of the polypeptide sequence of the present invention minus 6, and where "b" equals an integer between 7 and the number of
10 amino acids of the polypeptide sequence of the present invention; and where "a" is an integer smaller than "b" by at least 6.

The present invention also provides for the exclusion of any species of polypeptide fragments of the present invention specified by 5' and 3' positions or sub-genuses of polypeptides specified by size in amino acids as described above. Any number of fragments specified by 5' and
15 3' positions or by size in amino acids, as described above, may be excluded. Specifically excluded from the invention are the polypeptide fragments encoded by the preferentially excluded polynucleotide fragments described in Table IV, and in Tables Va and Vb. Table IV and Tables Va and Vb provide for the exclusion of polypeptides, independently from each other, in addition to those described elsewhere in the specification and is therefore, not meant as limiting description.

20 Functional definition

Preferred polypeptide fragments of the invention are isolated, purified or recombinant polypeptides comprising, consisting of, or consisting essentially of signal peptides, preferably signal peptides selected from the group consisting of SEQ ID Nos: 242-272 and 274-384, signal peptides encoded by sequences of SEQ ID Nos: 1-31 and 33-143 and those encoded by the clone inserts of
25 the deposited clone pool. Such polypeptides fragments are useful to design secretion vectors as described elsewhere in the application.

Other preferred polypeptide fragments of the invention are isolated, purified or recombinant polypeptides comprising, consisting of, or consisting essentially of mature proteins, preferably mature proteins selected from the group consisting of SEQ ID Nos: 242-272 and 274-384, mature
30 proteins encoded by sequences of SEQ ID Nos: 1-31 and 33-143 and those encoded by the clone inserts of the deposited clone pool.

Domains

Preferred polynucleotide fragments of the invention are domains of polypeptides of the invention. Such domains may eventually comprise linear or structural motifs and signatures
35 including, but not limited to, leucine zippers, helix-turn-helix motifs, post-translational modification sites such as glycosylation sites, ubiquitination sites, alpha helices, and beta sheets, signal sequences

encoding signal peptides which direct the secretion of the encoded proteins, sequences implicated in transcription regulation such as homeoboxes, acidic stretches, enzymatic active sites, substrate binding sites, and enzymatic cleavage sites. Such domains may present a particular biological activity such as DNA or RNA-binding, secretion of proteins, transcription regulation, enzymatic activity, substrate binding activity, etc...

A domain has a size generally comprised between 3 and 2000 amino acids. In preferred embodiment, domains comprise a number of amino acids that is any integer between 6 and 500. Domains may be synthesized using any methods known to those skilled in the art, including those disclosed herein, particularly in the section entitled "Preparation of the polypeptides of the invention". Methods for determining the amino acids which make up a domain with a particular biological activity include mutagenesis studies and assays to determine the biological activity to be tested.

Alternatively, the polypeptides of the invention may be scanned for motifs, domains and/or signatures in databases using any computer method known to those skilled in the art. Searchable databases include Prosite (Hofmann *et al.*, 1999; Bucher and Bairoch 1994), Pfam (Sonnhammer *et al.*, 1997; Henikoff *et al.*, 2000; Bateman *et al.*, 2000), Blocks (Henikoff *et al.*, 2000), Print (Attwood *et al.*, 1996), Prodom (Sonnhammer and Kahn, 1994; Corpet *et al.* 2000), Sbase (Pongor *et al.*, 1993; Murvai *et al.*, 2000), Smart (Schultz *et al.*, 1998), Dali/FSSP (Holm and Sander, 1996, 1997 and 1999), HSSP (Sander and Schneider 1991), CATH (Orengo *et al.*, 1997; Pearl *et al.*, 2000), SCOP (Murzin *et al.*, 1995; Lo Conte *et al.*, 2000), COG (Tatusov *et al.*, 1997 and 2000), specific family databases and derivatives thereof (Nevill-Manning *et al.*, 1998; Yona *et al.*, 1999; Attwood *et al.*, 2000), each of which disclosures are hereby incorporated by reference in their entireties. For a review on available databases, see issue 1 of volume 28 of Nucleic Acid Research (2000), which disclosure is hereby incorporated by reference in its entirety.

The polypeptides of SEQ ID NOs : 242-482 were screened for the presence of known structural or functional motifs or for the presence of signatures, small amino acid sequences that are well conserved amongst the members of a protein family. The search was conducted on the Pfam 5.5 database using HMMER-2.1.1 (for info see Sonnhammer et Durbin, <http://www.sanger.ac.uk/Pfam/>), on a Blocks Plus database containing Blocks version 12.0, Prints version 26.0, Pfam version 5.3, Prodom version 99.1, and Domo version 2.0 using emotif (for info see Nevill-Manning *et al.*, *PNAS*, 95, 5865-5871, (1998), <http://motif.stanford.edu/EMOTIF>) and on the Prosite 16.0 database using bla (Tatusov, R. L. & Koonin, E. V. CABIOS 10, No. 4) and pfscan (<http://www.isrec.isb-sib.ch/cgi-bin/man.cgi?section=1&topic=pfscan>). Some of these predicted domains are described in Table VI. For these polypeptides referred to by their sequence identification numbers (column entitled "Seq Id No"), Table VI gives the designation of the domain (column entitled "Designation of domain") according to the database of domains indicated in the column entitled "Database" and the positions of preferred fragments within these sequences (column

entitled "Positions of domains"). Each fragment is represented by a-b where a and b are the start and end positions respectively of a given preferred fragment on the full-length polypeptide. Preferred fragments are separated from each other by a comma. As used herein, the term "domain described in Table VI" refers to all the domains listed in Table VI for a given GENSET protein referred to by its sequence identification number in the first column. It should be noted that in Table VI, the first methionine encountered is designated as amino acid number 1, i.e., the leader sequence is not numbered negatively. In the appended sequence listing, the first amino acid of the mature protein resulting from cleavage of the signal peptide is designated as amino acid number 1 and the first amino acid of the signal peptide is designated with the appropriate negative number, in accordance with the regulations governing sequence listings.

Consequently, preferred polynucleotide fragments of the invention are domains of the polypeptides of SEQ ID Nos: 242-482. Therefore, the present invention encompasses isolated, purified, or recombinant polypeptides which consist of, consist essentially of, or comprise a contiguous span of at least 6, preferably at least 8 to 10, more preferably 12, 15, 20, 25, 30, 35, 40, 50, 60, 75, 100, 125, 150, 175, 200, 225, 250, 275, 300, 350, 400, 450 or 500 amino acids of a sequence selected from the group consisting of the sequences of SEQ ID Nos: 242-482, to the extent that a contiguous span of these lengths is consistent with the lengths of said selected sequence, where said contiguous span comprises at least 1, 2, 3, 5, or 10 amino acids positions of a domain described in Table VI of said selected sequence. The present invention also encompasses isolated, purified, or recombinant polypeptides comprising, consisting essentially of, or consisting of a contiguous span of at least 6, preferably at least 8 to 10, more preferably 12, 15, 20, 25, 30, 35, 40, 50, 60, 75, 100, 125, 150, 175, 200, 225, 250, 275, 300, 350, 400, 450 or 500 amino acids of a sequence selected from the group consisting of the sequences of SEQ ID Nos: 242-482, to the extent that a contiguous span of these lengths is consistent with the lengths of said selected sequence, where said contiguous span is a domain described in Table VI of said selected sequence. The present invention also encompasses isolated, purified, or recombinant polypeptides which comprise, consist of or consist essentially of a domain described in Table VI of a sequence selected from the group consisting of the sequences of SEQ ID Nos: 242-482.

Polypeptides of the present invention that are not specifically described in this table are not considered as not belonging to a domain. This is because they may still be not recognized as such by the particular algorithms used or not be included in the particular database searched. In fact, all fragments of the polypeptides of the present invention, at least 6 amino acids residues in length, are included in the present invention as being a domain. Amino acid residues comprising other domains may be determined by looking in other databases than the ones currently cited to establish Table VI. The domains of the present invention preferably comprises 6 to 200 amino acids (i.e. any integer between 6 and 200, inclusive) of a polypeptide of the present invention. Also, included in the present invention are domain fragments between the integers of 6 and the full length GENSET

sequence of the sequence listing. All combinations of sequences between the integers of 6 and the full-length sequence of a GENSET polypeptide are included. The domain fragments may be specified by either the number of contiguous amino acid residues (as a sub-genus) or by specific N-terminal and C-terminal positions (as species) as described above for the polypeptide fragments of the present invention. Any number of domain fragments of the present invention may also be excluded in the same manner.

Epitopes and Antibody Fusions:

A preferred embodiment of the present invention is directed to epitope-bearing polypeptides and epitope-bearing polypeptide fragments. These epitopes may be "antigenic epitopes" or both an "antigenic epitope" and an "immunogenic epitope". An "immunogenic epitope" is defined as a part of a protein that elicits an antibody response *in vivo* when the polypeptide is the immunogen. On the other hand, a region of polypeptide to which an antibody binds is defined as an "antigenic determinant" or "antigenic epitope." The number of immunogenic epitopes of a protein generally is less than the number of antigenic epitopes (*See, e.g., Geysen, et al., 1984*), which disclosure is hereby incorporated by reference in its entirety. It is particularly noted that although a particular epitope may not be immunogenic, it is nonetheless useful since antibodies can be made to both immunogenic and antigenic epitopes.

An epitope can comprise as few as 3 amino acids in a spatial conformation, which is unique to the epitope. Generally an epitope consists of at least 6 such amino acids, and more often at least 8-10 such amino acids. In preferred embodiment, antigenic epitopes comprise a number of amino acids that is any integer between 3 and 50. Fragments which function as epitopes may be produced by any conventional means (*See, e.g., Houghten, 1985*), also further described in U.S. Patent No. 4,631,21, which disclosures are hereby incorporated by reference in their entireties. Methods for determining the amino acids which make up an epitope include x-ray crystallography, 2-dimensional nuclear magnetic resonance, and epitope mapping, e.g., the Pepscan method described by Geysen *et al.* (1984); PCT Publication No. WO 84/03564; and PCT Publication No. WO 84/03506, which disclosures are hereby incorporated by reference in their entireties. Another example is the algorithm of Jameson and Wolf, (1988) (said reference incorporated by reference in its entirety). The Jameson-Wolf antigenic analysis, for example, may be performed using the computer program PROTEAN, using default parameters (Version 4.0 Windows, DNASTAR, Inc., 1228 South Park Street Madison, WI).

Antigenic epitopes predicted by the Jameson-Wolf algorithm for the polypeptides of SEQ ID Nos: 242-482 are presented in Table VII. For each GENSET polypeptide referred to by its sequence identification number in the column entitled "Seq Id No", a list of antigenic epitopes is given in the column entitled "Epitopes", each epitope being separated by a coma. Each fragment is represented by a-b where a and b are the start and end positions respectively of a given preferred

fragment. It should be noted that in Table VII, the first methionine encountered is designated as amino acid number 1, i.e; the leader sequence is not numbered negatively. In the appended sequence listing, the first amino acid of the mature protein resulting from cleavage of the signal peptide is designated as amino acid number 1 and the first amino acid of the signal peptide is
5 designated with the appropriate negative number, in accordance with the regulations governing sequence listings. As used herein, the term “epitope described in Table VII” refers to all preferred polynucleotide fragments described in the second column of Table VII for a GENSET polypeptide referred to by its sequence identification number in the first column. It is pointed out that the immunogenic epitopes listed in Table VII describe only amino acid residues comprising epitopes
10 predicted to have the highest degree of immunogenicity by a particular algorithm. Polypeptides of the present invention that are not specifically described as immunogenic are not considered non-antigenic. This is because they may still be antigenic *in vivo* but merely not recognized as such by the particular algorithm used. Alternatively, the polypeptides are most likely antigenic *in vitro* using methods such a phage display. Thus, listed in Table VII are the amino acid residues
15 comprising only preferred epitopes, not a complete list. In fact, all fragments of the polypeptides of the present invention, at least 6 amino acids residues in length, are included in the present invention as being useful as antigenic epitope. Amino acid residues comprising other immunogenic epitopes may be determined by algorithms similar to the Jameson-Wolf analysis or by *in vivo* testing for an antigenic response using the methods described herein or those known in the art.

20 Therefore, the present invention encompasses isolated, purified, or recombinant polypeptides which consist of, consist essentially of, or comprise a contiguous span of at least 6, preferably at least 8 to 10, more preferably 12, 15, 20, 25, 30, 35, 40, 50, 60, 75, 100, 125, 150, 175, 200, 225, 250, 275, 300, 350, 400, 450 or 500 amino acids of a sequence selected from the group consisting of the sequences of SEQ ID Nos: 242-482, to the extent that a contiguous span of these
25 lengths is consistent with the lengths of said selected sequence, where said contiguous span comprises at least 1, 2, 3, 5, or 10 amino acids positions of an epitope described in Table VII of said selected sequence. The present invention also encompasses isolated, purified, or recombinant polypeptides comprising, consisting essentially of, or consisting of a contiguous span of at least 6, preferably at least 8 to 10, more preferably 12, 15, 20, 25, 30, 35, 40, 50, 60, 75, 100, 125, 150, 175,
30 200, 225, 250, 275, 300, 350, 400, 450 or 500 amino acids of a sequence selected from the group consisting of the sequences of SEQ ID Nos: 242-482, to the extent that a contiguous span of these lengths is consistent with the lengths of said selected sequence, where said contiguous span is an epitope described in Table VII of said selected sequence. The present invention also encompasses isolated, purified, or recombinant polypeptides which comprise, consist of or consist essentially of
35 an epitope described in Table VII of a sequence selected from the group consisting of the sequences of SEQ ID Nos: 242-482.

The epitope-bearing fragments of the present invention preferably comprises 6 to 50 amino acids (i.e. any integer between 6 and 50, inclusive) of a polypeptide of the present invention. Also, included in the present invention are antigenic fragments between the integers of 6 and the full length GENSET sequence of the sequence listing. All combinations of sequences between the 5 integers of 6 and the full-length sequence of a GENSET polypeptide are included. The epitope-bearing fragments may be specified by either the number of contiguous amino acid residues (as a sub-genus) or by specific N-terminal and C-terminal positions (as species) as described above for the polypeptide fragments of the present invention. Any number of epitope-bearing fragments of the present invention may also be excluded in the same manner.

10 Antigenic epitopes are useful, for example, to raise antibodies, including monoclonal antibodies that specifically bind the epitope (See, Wilson *et al.*, 1984; and Sutcliffe, *et al.*, 1983), which disclosures are hereby incorporated by reference in their entireties. The antibodies are then used in various techniques such as diagnostic and tissue/cell identification techniques, as described herein, and in purification methods such as immunoaffinity chromatography.

15 An antibody or other compound that specifically binds to a polypeptide or polynucleotide of the invention is also said to "selectively recognize" the polypeptide or polynucleotide.

 Similarly, immunogenic epitopes can be used to induce antibodies according to methods well known in the art (See, Sutcliffe *et al.*, *supra*; Wilson *et al.*, *supra*; Chow *et al.*, (1985) and Bittle, *et al.*, (1985), which disclosures are hereby incorporated by reference in their entireties). A 20 preferred immunogenic epitope includes the natural GENSET protein. The immunogenic epitopes may be presented together with a carrier protein, such as an albumin, to an animal system (such as rabbit or mouse) or, if it is long enough (at least about 25 amino acids), without a carrier. However, immunogenic epitopes comprising as few as 8 to 10 amino acids have been shown to be sufficient to raise antibodies capable of binding to, at the very least, linear epitopes in a denatured polypeptide 25 (e.g., in Western blotting.).

 Epitope-bearing polypeptides of the present invention are used to induce antibodies according to methods well known in the art including, but not limited to, *in vivo* immunization, *in vitro* immunization, and phage display methods (See, e.g., Sutcliffe, *et al.*, *supra*; Wilson, *et al.*, *supra*, and Bittle, *et al.*, *supra*). If *in vivo* immunization is used, animals may be immunized with 30 free peptide; however, anti-peptide antibody titer may be boosted by coupling of the peptide to a macromolecular carrier, such as keyhole limpet hemacyanin (KLH) or tetanus toxoid. For instance, peptides containing cysteine residues may be coupled to a carrier using a linker such as -maleimidobenzoyl- N-hydroxysuccinimide ester (MBS), while other peptides may be coupled to carriers using a more general linking agent such as glutaraldehyde. Animals such as rabbits, rats 35 and mice are immunized with either free or carrier-coupled peptides, for instance, by intraperitoneal and/or intradermal injection of emulsions containing about 100 µgs of peptide or carrier protein and Freund's adjuvant. Several booster injections may be needed, for instance, at intervals of about two

weeks, to provide a useful titer of anti-peptide antibody, which can be detected, for example, by ELISA assay using free peptide adsorbed to a solid surface. The titer of anti-peptide antibodies in serum from an immunized animal may be increased by selection of anti-peptide antibodies, for instance, by adsorption to the peptide on a solid support and elution of the selected antibodies

5 according to methods well known in the art.

As one of skill in the art will appreciate, and discussed above, the polypeptides of the present invention comprising an immunogenic or antigenic epitope can be fused to heterologous polypeptide sequences. For example, the polypeptides of the present invention may be fused with the constant domain of immunoglobulins (IgA, IgE, IgG, IgM), or portions thereof (CH1, CH2,
10 CH3, any combination thereof including both entire domains and portions thereof) resulting in chimeric polypeptides. These fusion proteins facilitate purification, and show an increased half-life *in vivo*. This has been shown, *e.g.*, for chimeric proteins consisting of the first two domains of the human CD4-polypeptide and various domains of the constant regions of the heavy or light chains of mammalian immunoglobulins (*See, e.g.*, EPA 0,394,827; and Traunecker *et al.*, 1988), which
15 disclosures are hereby incorporated by reference in their entireties. Fusion proteins that have a disulfide-linked dimeric structure due to the IgG portion can also be more efficient in binding and neutralizing other molecules than monomeric polypeptides or fragments thereof alone (*See, e.g.*, Fountoulakis *et al.*, 1995), which disclosure is hereby incorporated by reference in its entirety. Nucleic acids encoding the above epitopes can also be recombined with a gene of interest as an
20 epitope tag to aid in detection and purification of the expressed polypeptide.

Additional fusion proteins of the invention may be generated through the techniques of gene-shuffling, motif-shuffling, exon-shuffling, or codon-shuffling (collectively referred to as "DNA shuffling"). DNA shuffling may be employed to modulate the activities of polypeptides of the present invention thereby effectively generating agonists and antagonists of the polypeptides.
25 *See, for example*, U.S. Patent Nos.: 5,605,793; 5,811,238; 5,834,252; 5,837,458; and Patten, *et al.*, (1997); Harayama, (1998); Hansson, *et al* (1999); and Lorenzo and Blasco, (1998). (Each of these documents are hereby incorporated by reference). In one embodiment, one or more components, motifs, sections, parts, domains, fragments, etc., of coding polynucleotides of the invention, or the polypeptides encoded thereby may be recombined with one or more components, motifs, sections,
30 parts, domains, fragments, etc. of one or more heterologous molecules.

The present invention further encompasses any combination of the polypeptide fragments listed in this section.

Antibodies:

Definitions

35 The present invention further relates to antibodies and T-cell antigen receptors (TCR), which specifically bind the polypeptides, and more specifically, the epitopes of the polypeptides of

the present invention. The antibodies of the present invention include IgG (including IgG1, IgG2, IgG3, and IgG4), IgA (including IgA1 and IgA2), IgD, IgE, or IgM, and IgY. The term "antibody" (Ab) refers to a polypeptide or group of polypeptides which are comprised of at least one binding domain, where a binding domain is formed from the folding of variable domains of an antibody molecule to form three-dimensional binding spaces with an internal surface shape and charge distribution complementary to the features of an antigenic determinant of an antigen, which allows an immunological reaction with the antigen. As used herein, the term "antibody" is meant to include whole antibodies, including single-chain whole antibodies, and antigen binding fragments thereof. In a preferred embodiment the antibodies are human antigen binding antibody fragments of the present invention include, but are not limited to, Fab, Fab' F(ab)2 and F(ab')2, Fd, single-chain Fvs (scFv), single-chain antibodies, disulfide-linked Fvs (sdFv) and fragments comprising either a V_L or V_H domain. The antibodies may be from any animal origin including birds and mammals. Preferably, the antibodies are human, murine, rabbit, goat, guinea pig, camel, horse, or chicken.

Antigen-binding antibody fragments, including single-chain antibodies, may comprise the variable region(s) alone or in combination with the entire or partial of the following: hinge region, CH1, CH2, and CH3 domains. Also included in the invention are any combinations of variable region(s) and hinge region, CH1, CH2, and CH3 domains. The present invention further includes chimeric, humanized, and human monoclonal and polyclonal antibodies, which specifically bind the polypeptides of the present invention. The present invention further includes antibodies that are anti-idiotypic to the antibodies of the present invention.

The antibodies of the present invention may be monospecific, bispecific, and trispecific or have greater multispecificity. Multispecific antibodies may be specific for different epitopes of a polypeptide of the present invention or may be specific for both a polypeptide of the present invention as well as for heterologous compositions, such as a heterologous polypeptide or solid support material. *See, e.g.*, WO 93/17715; WO 92/08802; WO 91/00360; WO 92/05793; Tutt, *et al.* (1991); US Patents 5,573,920, 4,474,893, 5,601,819, 4,714,681, 4,925,648; Kostelny *et al.* (1992), which disclosures are hereby incorporated by reference in their entireties.

Antibodies of the present invention may be described or specified in terms of the epitope(s) or epitope-bearing portion(s) of a polypeptide of the present invention, which are recognized or specifically bound by the antibody. The antibodies may specifically bind a complete protein encoded by a nucleic acid of the present invention, or a fragment thereof, particularly, in the case of secreted proteins the mature protein or the signal peptide. Therefore, the epitope(s) or epitope bearing polypeptide portion(s) may be specified as described herein, *e.g.*, by N-terminal and C-terminal positions, by size in contiguous amino acid residues, or otherwise described herein (including the sequence listing). Antibodies which specifically bind any epitope or polypeptide of the present invention may also be excluded as individual species. Therefore, the present invention includes

antibodies that specifically bind specified polypeptides of the present invention, and allows for the exclusion of the same.

Thus, another embodiment of the present invention is a purified or isolated antibody capable of specifically binding to a polypeptide comprising a sequence selected from the group consisting of the sequences of SEQ ID Nos: 242-482 and the sequences of the clone inserts of the deposited clone pool. In one aspect of this embodiment, the antibody is capable of binding to an epitope-containing polypeptide comprising at least 6 consecutive amino acids, preferably at least 8 to 10 consecutive amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 consecutive amino acids of a sequence selected from the group consisting of SEQ ID Nos: 242-482 and sequences of the clone inserts of the deposited clone pool.

Antibodies of the present invention may also be described or specified in terms of their cross-reactivity. Antibodies that do not specifically bind any other analog, ortholog, or homologue of the polypeptides of the present invention are included. Antibodies that do not bind polypeptides with less than 95%, less than 90%, less than 85%, less than 80%, less than 75%, less than 70%, less than 65%, less than 60%, less than 55%, and less than 50% identity (as calculated using methods known in the art and described herein, e.g., using FASTDB and the parameters set forth herein) to a polypeptide of the present invention are also included in the present invention. Further included in the present invention are antibodies, which only bind polypeptides encoded by polynucleotides, which hybridize to a polynucleotide of the present invention under stringent hybridization conditions (as described herein). Antibodies of the present invention may also be described or specified in terms of their binding affinity. Preferred binding affinities include those with a dissociation constant or K_d less than $5 \times 10^{-6}M$, $10^{-6}M$, $5 \times 10^{-7}M$, $10^{-7}M$, $5 \times 10^{-8}M$, $10^{-8}M$, $5 \times 10^{-9}M$, $10^{-9}M$, $5 \times 10^{-10}M$, $10^{-10}M$, $5 \times 10^{-11}M$, $10^{-11}M$, $5 \times 10^{-12}M$, $10^{-12}M$, $5 \times 10^{-13}M$, $10^{-13}M$, $5 \times 10^{-14}M$, $10^{-14}M$, $5 \times 10^{-15}M$, and $10^{-15}M$.

The invention also concerns a purified or isolated antibody capable of specifically binding to a mutated GENSET protein or to a fragment or variant thereof comprising an epitope of the mutated GENSET protein.

Preparation of antibodies

The antibodies of the present invention may be prepared by any suitable method known in the art. Some of these methods are described in more detail in the example entitled "Preparation of Antibody Compositions to ". For example, a polypeptide of the present invention or an antigenic fragment thereof can be administered to an animal in order to induce the production of sera containing "polyclonal antibodies". As used herein, the term "monoclonal antibody" is not limited to antibodies produced through hybridoma technology but it rather refers to an antibody that is derived from a single clone, including eukaryotic, prokaryotic, or phage clone, and not the method

by which it is produced. Monoclonal antibodies can be prepared using a wide variety of techniques known in the art including the use of hybridoma, recombinant, and phage display technology.

Hybridoma techniques include those known in the art (*See, e.g.,* Harlow *et al.* 1988; Hammerling, *et al.*, 1981). (Said references incorporated by reference in their entireties). Fab and
5 F(ab')₂ fragments may be produced, for example, from hybridoma-produced antibodies by proteolytic cleavage, using enzymes such as papain (to produce Fab fragments) or pepsin (to produce F(ab')₂ fragments).

Alternatively, antibodies of the present invention can be produced through the application of recombinant DNA technology or through synthetic chemistry using methods known in the art.
10 For example, the antibodies of the present invention can be prepared using various phage display methods known in the art. In phage display methods, functional antibody domains are displayed on the surface of a phage particle, which carries polynucleotide sequences encoding them. Phage with a desired binding property are selected from a repertoire or combinatorial antibody library (e.g. human or murine) by selecting directly with antigen, typically antigen bound or captured to a solid
15 surface or bead. Phage used in these methods are typically filamentous phage including fd and M13 with Fab, Fv or disulfide stabilized Fv antibody domains recombinantly fused to either the phage gene III or gene VIII protein. Examples of phage display methods that can be used to make the antibodies of the present invention include those disclosed in Brinkman *et al.* (1995); Ames, *et al.* (1995); Kettleborough, *et al.* (1994); Persic, *et al.* (1997); Burton *et al.* (1994); PCT/GB91/01134;
20 WO 90/02809; WO 91/10737; WO 92/01047; WO 92/18619; WO 93/11236; WO 95/15982; WO 95/20401; and US Patents 5,698,426, 5,223,409, 5,403,484, 5,580,717, 5,427,908, 5,750,753, 5,821,047, 5,571,698, 5,427,908, 5,516,637, 5,780,225, 5,658,727 and 5,733,743 (said references incorporated by reference in their entireties).

As described in the above references, after phage selection, the antibody coding regions
25 from the phage can be isolated and used to generate whole antibodies, including human antibodies, or any other desired antigen binding fragment, and expressed in any desired host including mammalian cells, insect cells, plant cells, yeast, and bacteria. For example, techniques to recombinantly produce Fab, Fab' F(ab)₂ and F(ab')₂ fragments can also be employed using methods known in the art such as those disclosed in WO 92/22324; Mullinax *et al.* (1992); and
30 Sawai *et al.* (1995); and Better *et al.* (1988) (said references incorporated by reference in their entireties).

Examples of techniques which can be used to produce single-chain Fvs and antibodies include those described in U.S. Patents 4,946,778 and 5,258,498; Huston *et al.* (1991); Shu *et al.* (1993); and Skerra *et al.* (1988), which disclosures are hereby incorporated by reference in their
35 entireties. For some uses, including *in vivo* use of antibodies in humans and *in vitro* detection assays, it may be preferable to use chimeric, humanized, or human antibodies. Methods for producing chimeric antibodies are known in the art. *See e.g.,* Morrison, (1985); Oi *et al.*, (1986);

Gillies *et al.* (1989); and US Patent 5,807,715, which disclosures are hereby incorporated by reference in their entireties. Antibodies can be humanized using a variety of techniques including CDR-grafting (EP 0 239 400; WO 91/09967; US Patent 5,530,101; and 5,585,089), veneering or resurfacing, (EP 0 592 106; EP 0 519 596; Padlan, 1991; Studnicka *et al.*, 1994; Roguska *et al.*, 1994), and chain shuffling (US Patent 5,565,332), which disclosures are hereby incorporated by reference in their entireties. Human antibodies can be made by a variety of methods known in the art including phage display methods described above. *See also*, US Patents 4,444,887, 4,716,111, 5,545,806, and 5,814,318; WO 98/46645; WO 98/50433; WO 98/24893; WO 96/34096; WO 96/33735; and WO 91/10741 (said references incorporated by reference in their entireties).

10 Further included in the present invention are antibodies recombinantly fused or chemically conjugated (including both covalently and non-covalently conjugations) to a polypeptide of the present invention. The antibodies may be specific for antigens other than polypeptides of the present invention. For example, antibodies of the present invention may be recombinantly fused or conjugated to molecules useful as labels in detection assays and effector molecules such as
15 heterologous polypeptides, drugs, or toxins. *See, e.g.*, WO 92/08495; WO 91/14438; WO 89/12624; US Patent 5,314,995; and EP 0 396 387, which disclosures are hereby incorporated by reference in their entireties. Fused antibodies may also be used to target the polypeptides of the present invention to particular cell types, either *in vitro* or *in vivo*, by fusing or conjugating the polypeptides of the present invention to antibodies specific for particular cell surface receptors. Antibodies fused
20 or conjugated to the polypeptides of the present invention may also be used in *in vitro* immunoassays and purification methods using methods known in the art (*See e.g.*, Harbor *et al. supra*; WO 93/21232; EP 0 439 095; Naramura, M. *et al.* 1994; US Patent 5,474,981; Gillies *et al.*, 1992; Fell *et al.*, 1991) (said references incorporated by reference in their entireties).

The present invention further includes compositions comprising the polypeptides of the
25 present invention fused or conjugated to antibody domains other than the variable regions. For example, the polypeptides of the present invention may be fused or conjugated to an antibody Fc region, or portion thereof. The antibody portion fused to a polypeptide of the present invention may comprise the hinge region, CH1 domain, CH2 domain, and CH3 domain or any combination of whole domains or portions thereof. The polypeptides of the present invention may be fused or
30 conjugated to the above antibody portions to increase the *in vivo* half-life of the polypeptides or for use in immunoassays using methods known in the art. The polypeptides may also be fused or conjugated to the above antibody portions to form multimers. For example, Fc portions fused to the polypeptides of the present invention can form dimers through disulfide bonding between the Fc portions. Higher multimeric forms can be made by fusing the polypeptides to portions of IgA and
35 IgM. Methods for fusing or conjugating the polypeptides of the present invention to antibody portions are known in the art. *See e.g.*, US Patents 5,336,603, 5,622,929, 5,359,046, 5,349,053, 5,447,851, 5,112,946; EP 0 307 434, EP 0 367 166; WO 96/04388, WO 91/06570; Ashkenazi *et al.*

(1991); Zheng *et al.* (1995); and Vil *et al.* (1992) (said references incorporated by reference in their entireties).

Non-human animals or mammals, whether wild-type or transgenic, which express a different species of GENSET than the one to which antibody binding is desired, and animals which do not express GENSET (i.e. a GENSET knock out animal as described herein) are particularly useful for preparing antibodies. GENSET knock out animals will recognize all or most of the exposed regions of a GENSET protein as foreign antigens, and therefore produce antibodies with a wider array of GENSET epitopes. Moreover, smaller polypeptides with only 10 to 30 amino acids may be useful in obtaining specific binding to any one of the GENSET proteins. In addition, the humoral immune system of animals which produce a species of GENSET that resembles the antigenic sequence will preferentially recognize the differences between the animal's native GENSET species and the antigen sequence, and produce antibodies to these unique sites in the antigen sequence. Such a technique will be particularly useful in obtaining antibodies that specifically bind to any one of the GENSET proteins.

The antibodies of the invention may be labeled by any one of the radioactive, fluorescent or enzymatic labels known in the art.

USES OF POLYNUCLEOTIDES

Uses of polynucleotides as reagents

The polynucleotides of the present invention, particularly those described in the "Oligonucleotide primers and probes" section, may be used as reagents in isolation procedures, diagnostic assays, and forensic procedures. For example, sequences from the GENSET polynucleotides of the invention may be detectably labeled and used as probes to isolate other sequences capable of hybridizing to them. In addition, sequences from the GENSET polynucleotides of the invention may be used to design PCR primers to be used in isolation, diagnostic, or forensic procedures.

In forensic analyses

PCR primers may be used in forensic analyses, such as the DNA fingerprinting techniques described below. Such analyses may utilize detectable probes or primers based on the sequences of the polynucleotides of the invention. Consequently, the present invention encompasses methods of identification of an individual using the polynucleotides of the invention in forensic analyses, wherein said method includes the steps of:

- a) obtaining a biological sample containing nucleic acid material from an individual;
- b) obtaining an identification pattern for this individual using the polynucleotides of the invention, particularly using GENSET primers and probes;
- c) comparing said identification pattern with a reference identification pattern; and

d) determining whether said identification pattern is identical to said reference identification pattern.

In one embodiment of this method, the identification pattern consists in sequences of amplicons obtained using GENSET primers as explained in the sections entitled "Forensic Matching by DNA Sequencing" and "Positive Identification by DNA Sequencing".

In another embodiment, the identification pattern consists in unique band or dot patterns obtained using any method described in the sections entitled "Southern Blot Forensic Identification", "Dot Blot Identification Procedure" and "Alternative "Fingerprint" Identification Technique".

10 *Forensic Matching by DNA Sequencing*

In one exemplary method, DNA samples are isolated from forensic specimens of, for example, hair, semen, blood or skin cells by conventional methods. A panel of PCR primers designed from different polynucleotides of the invention using any technique known to those skilled in the art including those described herein, is then utilized to amplify DNA of approximately 100-200 bases in length from the forensic specimen. Corresponding sequences are obtained from a test subject. Each of these identification DNAs is then sequenced using standard techniques, and a simple database comparison determines the differences, if any, between the sequences from the subject and those from the sample. Statistically significant differences between the suspect's DNA sequences and those from the sample conclusively prove a lack of identity. This lack of identity can be proven, for example, with only one sequence. Identity, on the other hand, should be demonstrated with a large number of sequences, all matching. Preferably, a minimum of 50 statistically identical sequences of 100 bases in length are used to prove identity between the suspect and the sample.

Positive Identification by DNA Sequencing

The "Forensic Matching by DNA Sequencing" technique described herein may also be used on a larger scale to provide a unique fingerprint-type identification of any individual. In this technique, primers are prepared from a large number of polynucleotides of the invention. Preferably, 20 to 50 different primers are used. These primers are used to obtain a corresponding number of PCR-generated DNA segments from the individual in question. Each of these DNA segments is sequenced. The database of sequences generated through this procedure uniquely identifies the individual from whom the sequences were obtained. The same panel of primers may then be used at any later time to absolutely correlate tissue or other biological specimen with that individual.

Southern Blot Forensic Identification

The "Positive Identification by DNA Sequencing" procedure described herein is repeated to obtain a panel of at least 10 amplified sequences from an individual and a specimen. Preferably, the panel contains at least 50 amplified sequences. More preferably, the panel contains 100 amplified sequences. In some embodiments, the panel contains 200 amplified sequences. This PCR-generated DNA is then digested with one or a combination of, preferably, four base specific restriction enzymes. Such enzymes are commercially available and known to those of skill in the art. After digestion, the resultant gene fragments are size separated in multiple duplicate wells on an agarose gel and transferred to nitrocellulose using Southern blotting techniques well known to those with skill in the art. For a review of Southern blotting see Davis *et al.* (1986), which disclosure is hereby incorporated by reference in its entirety.

A panel of probes based on the sequences of the polynucleotides of the invention, or fragments thereof of at least 10 bases, are radioactively or colorimetrically labeled using methods known in the art, such as nick translation or end labeling, and hybridized to the Southern blot using techniques known in the art. Preferably, the probe comprises at least 12, 15, or 17 consecutive nucleotides from the polynucleotide of the invention. More preferably, the probe comprises at least 20-30 consecutive nucleotides from the polynucleotide of the invention. In some embodiments, the probe comprises more than 30 nucleotides from the polynucleotide of the invention. In other embodiments, the probe comprises at least 40, at least 50, at least 75, at least 100, at least 150, or at least 200 consecutive nucleotides from the polynucleotide of the invention.

Preferably, at least 5 to 10 of these labeled probes are used, and more preferably at least about 20 or 30 are used to provide a unique pattern. The resultant bands appearing from the hybridization of a large sample of polynucleotide of the invention will be a unique identifier. Since the restriction enzyme cleavage will be different for every individual, the band pattern on the Southern blot will also be unique. Increasing the number of cDNA probes will provide a statistically higher level of confidence in the identification since there will be an increased number of sets of bands used for identification.

Dot Blot Identification Procedure

Another technique for identifying individuals using the polynucleotide sequences disclosed herein utilizes a dot blot hybridization technique.

Genomic DNA is isolated from nuclei of subject to be identified. Oligonucleotide probes of approximately 30 bp in length are synthesized that correspond to at least 10, preferably 50 sequences from the polynucleotide of the invention. The probes are used to hybridize to the genomic DNA through conditions known to those in the art. The oligonucleotides are end labeled with P³² using polynucleotide kinase (Pharmacia). Dot Blots are created by spotting the genomic DNA onto nitrocellulose or the like using a vacuum dot blot manifold (BioRad, Richmond California). The nitrocellulose filter containing the genomic sequences is baked or UV linked to the

filter, prehybridized and hybridized with labeled probe using techniques known in the art (Davis *et al.* 1986). The ^{32}P labeled DNA fragments are sequentially hybridized with successively stringent conditions to detect minimal differences between the 30 bp sequence and the DNA.

Tetramethylammonium chloride is useful for identifying clones containing small numbers of
5 nucleotide mismatches (Wood *et al.*, 1985). A unique pattern of dots distinguishes one individual from another individual.

Alternative "Fingerprint" Identification Technique

In a representative alternative fingerprinting procedure, the probes are derived from cDNAs. Preferably, a plurality of probes having sequences from different genes are used as follows.

10 Polynucleotides containing at least 10 consecutive bases from these sequences can be used as probes. Preferably, the probe comprises at least 12, 15, or 17 consecutive nucleotides from the polynucleotide of the invention. More preferably, the probe comprises at least 20-30 consecutive nucleotides from the polynucleotide of the invention. In some embodiments, the probe comprises more than 30 nucleotides from the polynucleotide of the invention. In other embodiments, the
15 probe comprises at least 40, at least 50, at least 75, at least 100, at least 150, or at least 200 consecutive nucleotides from the polynucleotide of the invention.

Oligonucleotides, generally 20-mers, are prepared from a large number, e.g. 50, 100, or 200, of polynucleotides of the invention using commercially available oligonucleotide services such as Genset, Paris, France. Cell samples from the test subject are processed for DNA using
20 techniques well known to those with skill in the art. The nucleic acid is digested with restriction enzymes such as EcoRI and XbaI. Following digestion, samples are applied to wells for electrophoresis. The procedure, as known in the art, may be modified to accommodate polyacrylamide electrophoresis, however in this example, samples containing 5 ug of DNA are loaded into wells and separated on 0.8% agarose gels. The gels are transferred onto nitrocellulose
25 using standard Southern blotting techniques.

10 ng of each of the oligonucleotides are pooled and end-labeled with P^{32} . The nitrocellulose is prehybridized with blocking solution and hybridized with the labeled probes. Following hybridization and washing, the nitrocellulose filter is exposed to X-Omat AR X-ray film. The resulting hybridization pattern will be unique for each individual.

30 It is additionally contemplated within this example that the number of probe sequences used can be varied for additional accuracy or clarity.

To find corresponding genomic DNA sequences

The GENSET cDNAs of the invention may also be used to clone sequences located upstream of the cDNAs of the invention on the corresponding genomic DNA. Such upstream
35 sequences may be capable of regulating gene expression, including promoter sequences, enhancer sequences, and other upstream sequences which influence transcription or translation levels. Once

identified and cloned, these upstream regulatory sequences may be used in expression vectors designed to direct the expression of an inserted gene in a desired spatial, temporal, developmental, or quantitative fashion.

Use of cDNAs or Fragments thereof to Clone Upstream Sequences from Genomic DNA

5 Sequences derived from polynucleotides of the inventions may be used to isolate the promoters of the corresponding genes using chromosome walking techniques. In one chromosome walking technique, which utilizes the GenomeWalker™ kit available from Clontech, five complete genomic DNA samples are each digested with a different restriction enzyme which has a 6 base recognition site and leaves a blunt end. Following digestion, oligonucleotide adapters are ligated to
10 each end of the resulting genomic DNA fragments.

For each of the five genomic DNA libraries, a first PCR reaction is performed according to the manufacturer's instructions (which are incorporated herein by reference) using an outer adaptor primer provided in the kit and an outer gene specific primer. The gene specific primer should be selected to be specific for the polynucleotide of the invention of interest and should have a melting
15 temperature, length, and location in the polynucleotide of the invention which is consistent with its use in PCR reactions. Each first PCR reaction contains 5ng of genomic DNA, 5 µl of 10X Tth reaction buffer, 0.2 mM of each dNTP, 0.2 µM each of outer adaptor primer and outer gene specific primer, 1.1 mM of Mg(OAc)₂, and 1 µl of the Tth polymerase 50X mix in a total volume of 50 µl. The reaction cycle for the first PCR reaction is as follows: 1 min at 94 degree Celsius / 2 sec at 94
20 degree Celsius, 3 min at 72 degree Celsius (7 cycles) / 2 sec at 94 degree Celsius, 3-min at 67 degree Celsius (32 cycles) / 5 min at 67 degree Celsius.

The product of the first PCR reaction is diluted and used as a template for a second PCR reaction according to the manufacturer's instructions using a pair of nested primers which are located internally on the amplicon resulting from the first PCR reaction. For example, 5 µl of the
25 reaction product of the first PCR reaction mixture may be diluted 180 times. Reactions are made in a 50 µl volume having a composition identical to that of the first PCR reaction except the nested primers are used. The first nested primer is specific for the adaptor, and is provided with the GenomeWalker™ kit. The second nested primer is specific for the particular polynucleotide of the invention for which the promoter is to be cloned and should have a melting temperature, length, and
30 location in the polynucleotide of the invention which is consistent with its use in PCR reactions. The reaction parameters of the second PCR reaction are as follows: 1 min at 94 degree Celsius / 2 sec at 94 degree Celsius, 3 min at 72 degree Celsius (6 cycles) / 2 sec at 94 degree Celsius, 3 min at 67 degree Celsius (25 cycles) / 5 min at 67 degree Celsius

The product of the second PCR reaction is purified, cloned, and sequenced using standard
35 techniques. Alternatively, two or more human genomic DNA libraries can be constructed by using two or more restriction enzymes. The digested genomic DNA is cloned into vectors which can be

converted into single stranded, circular, or linear DNA. A biotinylated oligonucleotide comprising at least 15 nucleotides from the polynucleotide of the invention sequence is hybridized to the single stranded DNA. Hybrids between the biotinylated oligonucleotide and the single stranded DNA containing the polynucleotide of the invention sequence are isolated as described herein.

- 5 Thereafter, the single stranded DNA containing the polynucleotide of the invention sequence is released from the beads and converted into double stranded DNA using a primer specific for the polynucleotide of the invention sequence or a primer corresponding to a sequence included in the cloning vector. The resulting double stranded DNA is transformed into bacteria. DNAs containing the GENSET polynucleotide sequences are identified by colony PCR or colony hybridization.

10 *Identification of Promoters in Cloned Upstream Sequences*

Once the upstream genomic sequences have been cloned and sequenced as described above, prospective promoters and transcription start sites within the upstream sequences may be identified by comparing the sequences upstream of the polynucleotides of the inventions with databases containing known transcription start sites, transcription factor binding sites, or promoter sequences.

- 15 In addition, promoters in the upstream sequences may be identified using promoter reporter vectors as follows. The expression of the reporter gene will be detected when placed under the control of regulatory active polynucleotide fragments or variants of the GENSET promoter region located upstream of the first exon of the GENSET gene. Suitable promoter reporter vectors, into which the GENSET promoter sequences may be cloned include pSEAP-Basic, pSEAP-Enhancer, 20 p β gal-Basic, p β gal-Enhancer, or pEGFP-1 Promoter Reporter vectors available from Clontech, or pGL2-basic or pGL3-basic promoterless luciferase reporter gene vector from Promega. Briefly, each of these promoter reporter vectors include multiple cloning sites positioned upstream of a reporter gene encoding a readily assayable protein such as secreted alkaline phosphatase, luciferase, beta-galactosidase, or green fluorescent protein. The sequences upstream the GENSET coding 25 region are inserted into the cloning sites upstream of the reporter gene in both orientations and introduced into an appropriate host cell. The level of reporter protein is assayed and compared to the level obtained from a vector which lacks an insert in the cloning site. The presence of an elevated expression level in the vector containing the insert with respect to the control vector indicates the presence of a promoter in the insert. If necessary, the upstream sequences can be 30 cloned into vectors which contain an enhancer for increasing transcription levels from weak promoter sequences. A significant level of expression above that observed with the vector lacking an insert indicates that a promoter sequence is present in the inserted upstream sequence.

- Promoter sequence within the upstream genomic DNA may be further defined by site directed mutagenesis, linker scanning analysis, or other techniques familiar to those skilled in the 35 art. For example, the boundaries of promoters may be further investigated by constructing nested 5' and/or 3' deletions in the upstream DNA using conventional techniques such as Exonuclease III or

appropriate restriction endonuclease digestion. The resulting deletion fragments can be inserted into the promoter reporter vector to determine whether the deletion has increased, reduced or illuminated promoter activity, such as described, for example, by Coles *et al.* (1998), the disclosure of which is incorporated herein by reference in its entirety. In this way, the boundaries of the promoters may be defined. If desired, potential individual regulatory sites within the promoter may be identified using site directed mutagenesis or linker scanning to obliterate potential transcription factor binding sites within the promoter individually or in combination. The effects of these mutations on transcription levels may be determined by inserting the mutations into cloning sites in promoter reporter vectors. This type of assay is well known to those skilled in the art and is described in WO 97/17359, US Patent No. 5,374,544; EP 582 796; US Patent No. 5,698,389; US 5,643,746; US Patent No. 5,502,176; and US Patent 5,266,488; the disclosures of which are incorporated by reference herein in their entirety.

The strength and the specificity of the promoter of each GENSET gene can be assessed through the expression levels of a detectable polynucleotide operably linked to the GENSET promoter in different types of cells and tissues. The detectable polynucleotide may be either a polynucleotide that specifically hybridizes with a predefined oligonucleotide probe, or a polynucleotide encoding a detectable protein, including a GENSET polypeptide or a fragment or a variant thereof. This type of assay is well known to those skilled in the art and is described in US Patent No. 5,502,176; and US Patent No. 5,266,488; the disclosures of which are incorporated by reference herein in their entirety. Some of the methods are discussed in more detail elsewhere in the application.

The promoters and other regulatory sequences located upstream of the polynucleotides of the inventions may be used to design expression vectors capable of directing the expression of an inserted gene in a desired spatial, temporal, developmental, or quantitative manner. A promoter capable of directing the desired spatial, temporal, developmental, and quantitative patterns may be selected using the results of the expression analysis described herein. For example, if a promoter which confers a high level of expression in muscle is desired, the promoter sequence upstream of a polynucleotide of the invention derived from an mRNA which is expressed at a high level in muscle may be used in the expression vector. Such vectors are described in more detail elsewhere in the application.

Preferably, the desired promoter is placed near multiple restriction sites to facilitate the cloning of the desired insert downstream of the promoter, such that the promoter is able to drive expression of the inserted gene. The promoter may be inserted in conventional nucleic acid backbones designed for extrachromosomal replication, integration into the host chromosomes or transient expression. Suitable backbones for the present expression vectors include retroviral backbones, backbones from eukaryotic episomes such as SV40 or Bovine Papilloma Virus, backbones from bacterial episomes, or artificial chromosomes.

Preferably, the expression vectors also include a polyA signal downstream of the multiple restriction sites for directing the polyadenylation of mRNA transcribed from the gene inserted into the expression vector.

To find similar sequences

5 Polynucleotides of the invention may be used to isolate and/or purify nucleic acids similar thereto using any methods well known to those skilled in the art including the techniques based on hybridization or on amplification described in this section. These methods may be used to obtain the genomic DNAs which encode the mRNAs from which the GENSET cDNAs are derived, mRNAs corresponding to GENSET cDNAs, or nucleic acids which are homologous to GENSET cDNAs or
10 fragments thereof, such as variants, species homologues or orthologs. Thus, a plurality of cDNAs similar to GENSET polynucleotides may be provided as cDNA libraries for subsequent evaluation of the encoded proteins or use in diagnostic assays as described herein. cDNAs prepared by any method described therein may be subsequently engineered to obtain nucleic acids which include desired fragments of the cDNA using conventional techniques such as subcloning, PCR, or *in vitro*
15 oligonucleotide synthesis. For example, nucleic acids which include only the coding sequences may be obtained using techniques known to those skilled in the art. Similarly, nucleic acids containing any other desired fragment of the coding sequences for the encoded protein may be obtained.

Indeed, cDNAs of the present invention or fragments thereof may be used to isolate nucleic
20 acids similar to cDNAs from a cDNA library or a genomic DNA library. Such cDNA libraries or genomic DNA libraries may be obtained from a commercial source or made using techniques familiar to those skilled in the art such as those described in PCT publication WO 00/37491, which disclosure is hereby incorporated by reference in its entirety. Examples of methods for obtaining nucleic acids similar to GENSET polynucleotides are described below.

25 *Hybridization-based methods*

Techniques for identifying cDNA clones in a cDNA library which hybridize to a given probe sequence are disclosed in Sambrook *et al.*, (1989) and in Hames and Higgins (1985), the disclosures of which are incorporated herein by reference in their entireties. The same techniques may be used to isolate genomic DNAs.

30 Briefly, cDNA or genomic DNA clones which hybridize to the detectable probe are identified and isolated for further manipulation as follows. Any polynucleotide fragment of the invention may be used as a probe, in particular those defined in the "Oligonucleotide primers and probes" section. A probe comprising at least 10 consecutive nucleotides from a GENSET cDNA or fragment thereof is labeled with a detectable label such as a radioisotope or a fluorescent molecule.
35 Preferably, the probe comprises at least 12, 15, or 17 consecutive nucleotides from the cDNA or fragment thereof. More preferably, the probe comprises 20 to 30 consecutive nucleotides from the

cDNA or fragment thereof. In some embodiments, the probe comprises more than 30 nucleotides from the cDNA or fragment thereof.

Techniques for labeling the probe are well known and include phosphorylation with polynucleotide kinase, nick translation, *in vitro* transcription, and non radioactive techniques. The cDNAs or genomic DNAs in the library are transferred to a nitrocellulose or nylon filter and denatured. After blocking of non specific sites, the filter is incubated with the labeled probe for an amount of time sufficient to allow binding of the probe to cDNAs or genomic DNAs containing a sequence capable of hybridizing thereto.

By varying the stringency of the hybridization conditions used to identify cDNAs or genomic DNAs which hybridize to the detectable probe, cDNAs or genomic DNAs having different levels of identity to the probe can be identified and isolated as described below.

Stringent conditions

"Stringent hybridization conditions" are defined as conditions in which only nucleic acids having a high level of identity to the probe are able to hybridize to said probe. These conditions may be calculated as follows:

For probes between 14 and 70 nucleotides in length the melting temperature (T_m) is calculated using the formula: $T_m = 81.5 + 16.6(\log(Na^+)) + 0.41(\text{fraction G+C}) - (600/N)$ where N is the length of the probe.

If the hybridization is carried out in a solution containing formamide, the melting temperature may be calculated using the equation: $T_m = 81.5 + 16.6(\log(Na^+)) + 0.41(\text{fraction G+C}) - (0.63\% \text{ formamide}) - (600/N)$ where N is the length of the probe.

Prehybridization may be carried out in 6X SSC, 5X Denhardt's reagent, 0.5% SDS, 100 μ g denatured fragmented salmon sperm DNA or 6X SSC, 5X Denhardt's reagent, 0.5% SDS, 100 μ g denatured fragmented salmon sperm DNA, 50% formamide. The formulas for SSC and Denhardt's solutions are listed in Sambrook *et al.*, 1986.

Hybridization is conducted by adding the detectable probe to the prehybridization solutions listed above. Where the probe comprises double stranded DNA, it is denatured before addition to the hybridization solution. The filter is contacted with the hybridization solution for a sufficient period of time to allow the probe to hybridize to nucleic acids containing sequences complementary thereto or homologous thereto. For probes over 200 nucleotides in length, the hybridization may be carried out at 15-25°C below the T_m . For shorter probes, such as oligonucleotide probes, the hybridization may be conducted at 15-25°C below the T_m . Preferably, for hybridizations in 6X SSC, the hybridization is conducted at approximately 68°C. Preferably, for hybridizations in 50% formamide containing solutions, the hybridization is conducted at approximately 42°C.

Following hybridization, the filter is washed in 2X SSC, 0.1% SDS at room temperature for 15 minutes. The filter is then washed with 0.1X SSC, 0.5% SDS at room temperature for 30

minutes to 1 hour. Thereafter, the solution is washed at the hybridization temperature in 0.1X SSC, 0.5% SDS. A final wash is conducted in 0.1X SSC at room temperature.

Nucleic acids which have hybridized to the probe are identified by autoradiography or other conventional techniques.

5 Low and moderate conditions

Changes in the stringency of hybridization and signal detection are primarily accomplished through the manipulation of formamide concentration (lower percentages of formamide result in lowered stringency); salt conditions, or temperature. The above procedure may thus be modified to identify nucleic acids having decreasing levels of identity to the probe sequence. For example, the hybridization temperature may be decreased in increments of 5°C from 68°C to 42°C in a hybridization buffer having a sodium concentration of approximately 1M. Following hybridization, the filter may be washed with 2X SSC, 0.5% SDS at the temperature of hybridization. These conditions are considered to be "moderate" conditions above 50°C and "low" conditions below 50°C. Alternatively, the hybridization may be carried out in buffers, such as 6X SSC, containing formamide at a temperature of 42°C. In this case, the concentration of formamide in the hybridization buffer may be reduced in 5% increments from 50% to 0% to identify clones having decreasing levels of identity to the probe. Following hybridization, the filter may be washed with 6X SSC, 0.5% SDS at 50°C. These conditions are considered to be "moderate" conditions above 25% formamide and "low" conditions below 25% formamide. cDNAs or genomic DNAs which have hybridized to the probe are identified by autoradiography or other conventional techniques.

Note that variations in the above conditions may be accomplished through the inclusion and/or substitution of alternate blocking reagents used to suppress background in hybridization experiments. Typical blocking reagents include Denhardt's reagent, BLOTTO, heparin, denatured salmon sperm DNA, and commercially available proprietary formulations. The inclusion of specific blocking reagents may require modification of the hybridization conditions described above, due to problems with compatibility.

Consequently, the present invention encompasses methods of isolating nucleic acids similar to the polynucleotides of the invention, comprising the steps of:

- a) contacting a collection of cDNA or genomic DNA molecules with a detectable probe comprising at least 12, 15, 18, 20, 23, 25, 28, 30, 35, 40 or 50 consecutive nucleotides of a sequence selected from the group consisting of the sequences of SEQ ID Nos: 1-241, the sequences of clones inserts of the deposited clone pool and sequences complementary thereto under stringent, moderate or low conditions which permit said probe to hybridize to at least a cDNA or genomic DNA molecule in said collection;
- b) identifying said cDNA or genomic DNA molecule which hybridizes to said detectable probe; and

c) isolating said cDNA or genomic DNA molecule which hybridized to said probe.

PCR-based methods

In addition to the above described methods, other protocols are available to obtain homologous cDNAs using GENSET cDNA of the present invention or fragment thereof as outlined
5 in the following paragraphs.

cDNAs may be prepared by obtaining mRNA from the tissue, cell, or organism of interest using mRNA preparation procedures utilizing polyA selection procedures or other techniques known to those skilled in the art. A first primer capable of hybridizing to the polyA tail of the mRNA is hybridized to the mRNA and a reverse transcription reaction is performed to generate a
10 first cDNA strand.

The term "capable of hybridizing to the polyA tail of said mRNA" refers to and embraces all primers containing stretches of thymidine residues, so-called oligo(dT) primers, that hybridize to the 3' end of eukaryotic poly(A)+ mRNAs to prime the synthesis of a first cDNA strand. Techniques for generating said oligo (dT) primers and hybridizing them to mRNA to subsequently
15 prime the reverse transcription of said hybridized mRNA to generate a first cDNA strand are well known to those skilled in the art and are described in Current Protocols in Molecular Biology, John Wiley and Sons, Inc. 1997 and Sambrook *et al.*, 1989. Preferably, said oligo (dT) primers are present in a large excess in order to allow the hybridization of all mRNA 3'ends to at least one oligo (dT) molecule. The priming and reverse transcription steps are preferably performed between 37°C
20 and 55°C depending on the type of reverse transcriptase used. Preferred oligo(dT) primers for priming reverse transcription of mRNAs are oligonucleotides containing a stretch of thymidine residues of sufficient length to hybridize specifically to the polyA tail of mRNAs, preferably of 12 to 18 thymidine residues in length. More preferably, such oligo(T) primers comprise an additional sequence upstream of the poly(dT) stretch in order to allow the addition of a given sequence to the
25 5'end of all first cDNA strands which may then be used to facilitate subsequent manipulation of the cDNA. Preferably, this added sequence is 8 to 60 residues in length. For instance, the addition of a restriction site in 5' of cDNAs facilitates subcloning of the obtained cDNA. Alternatively, such an added 5'end may also be used to design primers of PCR to specifically amplify cDNA clones of interest.

30 The first cDNA strand is then hybridized to a second primer. Any polynucleotide fragment of the invention may be used, and in particular those described in the "Oligonucleotide primers and probes" section. This second primer contains at least 10 consecutive nucleotides of a polynucleotide of the invention. Preferably, the primer comprises at least 10, 12, 15, 17, 18, 20, 23, 25, or 28 consecutive nucleotides of a polynucleotide of the invention. In some embodiments, the primer
35 comprises more than 30 nucleotides of a polynucleotide of the invention. If it is desired to obtain cDNAs containing the full protein coding sequence, including the authentic translation initiation

site, the second primer used contains sequences located upstream of the translation initiation site. The second primer is extended to generate a second cDNA strand complementary to the first cDNA strand. Alternatively, RT-PCR may be performed as described above using primers from both ends of the cDNA to be obtained.

- 5 The double stranded cDNAs made using the methods described above are isolated and cloned. The cDNAs may be cloned into vectors such as plasmids or viral vectors capable of replicating in an appropriate host cell. For example, the host cell may be a bacterial, mammalian, avian, or insect cell.

Techniques for isolating mRNA, reverse transcribing a primer hybridized to mRNA to
10 generate a first cDNA strand, extending a primer to make a second cDNA strand complementary to the first cDNA strand, isolating the double stranded cDNA and cloning the double stranded cDNA are well known to those skilled in the art and are described in *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc. 1997 and Sambrook *et al.*, 1989.

Consequently, the present invention encompasses methods of making cDNAs. In a first
15 embodiment, the method of making a cDNA comprises the steps of

- a) contacting a collection of mRNA molecules from human cells with a primer comprising at least 12, 15, 18, 20, 23, 25, 28, 30, 35, 40, or 50 consecutive nucleotides of a sequence selected from the group consisting of the sequences complementary to SEQ ID Nos: 1-241 and sequences complementary to a clone insert of the deposited clone pool;
- 20 b) hybridizing said primer to an mRNA in said collection;
- c) reverse transcribing said hybridized primer to make a first cDNA strand from said mRNA;
- d) making a second cDNA strand complementary to said first cDNA strand; and
- e) isolating the resulting cDNA comprising said first cDNA strand and said second cDNA
25 strand.

Another embodiment of the present invention is a purified cDNA obtainable by the method of the preceding paragraph. In one aspect of this embodiment, the cDNA encodes at least a portion of a human polypeptide.

- In a second embodiment, the method of making a cDNA comprises the steps of
- 30 a) contacting a collection of mRNA molecules from human cells with a first primer capable of hybridizing to the polyA tail of said mRNA;
 - b) hybridizing said first primer to said polyA tail;
 - c) reverse transcribing said mRNA to make a first cDNA strand;
 - d) making a second cDNA strand complementary to said first cDNA strand using at least
35 one primer comprising at least 12, 15, 18, 20, 23, 25, 28, 30, 35, 40, or 50 consecutive nucleotides of a sequence selected from the group consisting of SEQ ID Nos: 1-241 and sequences of clone inserts of the deposited clone pool; and

e) isolating the resulting cDNA comprising said first cDNA strand and said second cDNA strand.

In another aspect of this method the second cDNA strand is made by

a) contacting said first cDNA strand with a second primer comprising at least 12, 15, 18, 20, 23, 25, 28, 30, 35, 40, or 50 consecutive nucleotides of a sequence selected from the group consisting of SEQ ID Nos: 1-241 and sequences of clone inserts of the deposited clone pool, and a third primer which sequence is fully included within the sequence of said first primer;

b) performing a first polymerase chain reaction with said second and third primers to generate a first PCR product;

c) contacting said first PCR product with a fourth primer, comprising at least 12, 15, 18, 20, 23, 25, 28, 30, 35, 40, or 50 consecutive nucleotides of said sequence selected from the group consisting of SEQ ID Nos: 1-241 and sequences of clone inserts of the deposited clone pool, and a fifth primer, which sequence is fully included within the sequence of said third primer, wherein said fourth and fifth hybridize to sequences within said first PCR product; and

d) performing a second polymerase chain reaction, thereby generating a second PCR product.

Alternatively, the second cDNA strand may be made by contacting said first cDNA strand with a second primer comprising at least 12, 15, 18, 20, 23, 25, 28, 30, 35, 40, or 50 consecutive nucleotides of a sequence selected from the group consisting of SEQ ID Nos: 1-241 and sequences of clone inserts of the deposited clone pool, and a third primer which sequence is fully included within the sequence of said first primer and performing a polymerase chain reaction with said second and third primers to generate said second cDNA strand.

Alternatively, the second cDNA strand may be made by

a) contacting said first cDNA strand with a second primer comprising at least 12, 15, 18, 20, 23, 25, 28, 30, 35, 40, or 50 consecutive nucleotides of a sequence selected from the group consisting of SEQ ID Nos: 1-241 and sequences of clone inserts of the deposited clone pool;

b) hybridizing said second primer to said first strand cDNA; and

c) extending said hybridized second primer to generate said second cDNA strand.

Another embodiment of the present invention is a purified cDNA obtainable by a method of making a cDNA of the invention. In one aspect of this embodiment, said cDNA encodes at least a portion of a human polypeptide.

Other protocols

Alternatively, other procedures may be used for obtaining homologous cDNAs. In one approach, cDNAs are prepared from mRNA and cloned into double stranded phagemids as follows. The cDNA library in the double stranded phagemids is then rendered single stranded by treatment with an endonuclease, such as the Gene II product of the phage F1 and an exonuclease (Chang *et*

al., 1993, which disclosure is hereby incorporated by reference in its entirety). A biotinylated oligonucleotide comprising the sequence of a fragment of a known GENSET cDNA, genomic DNA or fragment thereof is hybridized to the single stranded phagemids. Preferably, the fragment comprises at least 10, 12, 15, 17, 18, 20, 23, 25, or 28 consecutive nucleotides of a sequence
5 selected from the group consisting of the sequences of SEQ ID Nos: 1-241 and sequences of clone inserts of the deposited clone pool.

Hybrids between the biotinylated oligonucleotide and phagemids are isolated by incubating the hybrids with streptavidin coated paramagnetic beads and retrieving the beads with a magnet (Fry *et al.*, 1992, which disclosure is hereby incorporated by reference in its entirety). Thereafter,
10 the resulting phagemids are released from the beads and converted into double stranded DNA using a primer specific for the GENSET cDNA or fragment used to design the biotinylated oligonucleotide. Alternatively, protocols such as the Gene Trapper kit (Gibco BRL), which disclosure is which disclosure is hereby incorporated by reference in its entirety, may be used. The resulting double stranded DNA is transformed into bacteria. Homologous cDNAs to the GENSET
15 cDNA or fragment thereof sequence are identified by colony PCR or colony hybridization.

As a chromosome marker

Chromosomal localization of the cDNA of the present invention were determined using information from public and proprietary databases. Table VIII lists the putative chromosomal location of the polynucleotides of the present invention. Column one lists the sequence identification
20 number with the corresponding chromosomal location listed in column two. Thus, the present invention also relates to methods and compositions using the chromosomal location of the polynucleotides of the invention to construct a human high resolution map or to identify a given chromosome in a sample using any techniques known to those skilled in the art including those disclosed below.

25 GENSET polynucleotides may also be mapped to their chromosomal locations using any methods or techniques known to those skilled in the art including radiation hybrid (RH) mapping, PCR-based mapping and Fluorescence in situ hybridization (FISH) mapping described below.

Radiation hybrid mapping

Radiation hybrid (RH) mapping is a somatic cell genetic approach that can be used for high
30 resolution mapping of the human genome. In this approach, cell lines containing one or more human chromosomes are lethally irradiated, breaking each chromosome into fragments whose size depends on the radiation dose. These fragments are rescued by fusion with cultured rodent cells, yielding subclones containing different fragments of the human genome. This technique is described by Benham *et al.* (1989) and Cox *et al.*, (1990), which disclosures are hereby
35 incorporated by reference in their entireties. The random and independent nature of the subclones permits efficient mapping of any human genome marker. Human DNA isolated from a panel of 80-

100 cell lines provides a mapping reagent for ordering GENSET cDNAs or genomic DNAs. In this approach, the frequency of breakage between markers is used to measure distance, allowing construction of fine resolution maps as has been done using conventional ESTs (Schuler *et al.*, 1996), which disclosure is hereby incorporated by reference in its entirety.

- 5 RH mapping has been used to generate a high-resolution whole genome radiation hybrid map of human chromosome 17q22-q25.3 across the genes for growth hormone (GH) and thymidine kinase (TK) (Foster *et al.*, 1996), the region surrounding the Gorlin syndrome gene (Obermayr *et al.*, 1996), 60 loci covering the entire short arm of chromosome 12 (Raeymaekers *et al.*, 1995), the region of human chromosome 22 containing the neurofibromatosis type 2 locus (Frazer *et al.*, 1992) and 13 loci on the long arm of chromosome 5 (Warrington *et al.*, 1991), which disclosures are hereby incorporated by reference in their entireties.

Mapping of cDNAs to Human Chromosomes using PCR techniques

- GENSET cDNAs and genomic DNAs may be assigned to human chromosomes using PCR based methodologies. In such approaches, oligonucleotide primer pairs are designed from the cDNA sequence to minimize the chance of amplifying through an intron. Preferably, the oligonucleotide primers are 18-23 bp in length and are designed for PCR amplification. The creation of PCR primers from known sequences is well known to those with skill in the art. For a review of PCR technology see Erlich (1992), which disclosure is hereby incorporated by reference in its entirety.

- 20 The primers are used in polymerase chain reactions (PCR) to amplify templates from total human genomic DNA. PCR conditions are as follows: 60 ng of genomic DNA is used as a template for PCR with 80 ng of each oligonucleotide primer, 0.6 unit of Taq polymerase, and 1 uCi of a ³²P-labeled deoxycytidine triphosphate. The PCR is performed in a microplate thermocycler (Techne) under the following conditions: 30 cycles of 94 degree Celsius, 1.4 min; 55 degree Celsius, 2 min; and 72 degree Celsius, 2 min; with a final extension at 72 degree Celsius for 10 min. The amplified products are analyzed on a 6% polyacrylamide sequencing gel and visualized by autoradiography. If the length of the resulting PCR product is identical to the distance between the ends of the primer sequences in the cDNA from which the primers are derived, then the PCR reaction is repeated with DNA templates from two panels of human-rodent somatic cell hybrids, BIOS PCRable DNA (BIOS Corporation) and NIGMS Human-Rodent Somatic Cell Hybrid Mapping Panel Number 1 (NIGMS, Camden, NJ).

- PCR is used to screen a series of somatic cell hybrid cell lines containing defined sets of human chromosomes for the presence of a given cDNA or genomic DNA. DNA is isolated from the somatic hybrids and used as starting templates for PCR reactions using the primer pairs from the GENSET cDNAs or genomic DNAs. Only those somatic cell hybrids with chromosomes containing the human gene corresponding to the GENSET cDNA or genomic DNA will yield an

amplified fragment. The GENSET cDNAs or genomic DNAs are assigned to a chromosome by analysis of the segregation pattern of PCR products from the somatic hybrid DNA templates. The single human chromosome present in all cell hybrids that give rise to an amplified fragment is the chromosome containing that GENSET cDNA or genomic DNA. For a review of techniques and
5 analysis of results from somatic cell gene mapping experiments, see Ledbetter *et al.*, (1990), which disclosure is hereby incorporated by reference in its entirety.

Mapping of cDNAs to Chromosomes Using Fluorescence in situ Hybridization

Fluorescence *in situ* hybridization allows the GENSET cDNA or genomic DNA to be mapped to a particular location on a given chromosome. The chromosomes to be used for
10 fluorescence in situ hybridization techniques may be obtained from a variety of sources including cell cultures, tissues, or whole blood.

In a preferred embodiment, chromosomal localization of a GENSET cDNA or genomic DNA is obtained by FISH as described by Cherif *et al.* (1990), which disclosure is hereby incorporated by reference in its entirety. Metaphase chromosomes are prepared from
15 phytohemagglutinin (PHA)-stimulated blood cell donors. PHA-stimulated lymphocytes from healthy males are cultured for 72 h in RPMI-1640 medium. For synchronization, methotrexate (10 μ M) is added for 17 h, followed by addition of 5-bromodeoxyuridine (5-BudR, 0.1 mM) for 6 h. Colcemid (1 μ g/ml) is added for the last 15 min before harvesting the cells. Cells are collected,
20 washed in RPMI, incubated with a hypotonic solution of KCl (75 mM) at 37 degree Celsius for 15 min and fixed in three changes of methanol:acetic acid (3:1). The cell suspension is dropped onto a glass slide and air dried. The GENSET cDNA or genomic DNA is labeled with biotin-16 dUTP by nick translation according to the manufacturer's instructions (Bethesda Research Laboratories, Bethesda, MD), purified using a Sephadex G-50 column (Pharmacia, Upssala, Sweden) and precipitated. Just prior to hybridization, the DNA pellet is dissolved in hybridization buffer (50%
25 formamide, 2 X SSC, 10% dextran sulfate, 1 mg/ml sonicated salmon sperm DNA, pH 7) and the probe is denatured at 70 degree Celsius for 5-10 min.

Slides kept at -20 degree Celsius are treated for 1 h at 37 degree Celsius with RNase A (100 μ g/ml), rinsed three times in 2 X SSC and dehydrated in an ethanol series. Chromosome preparations are denatured in 70% formamide, 2 X SSC for 2 min at 70 degree Celsius, then
30 dehydrated at 4 degree Celsius. The slides are treated with proteinase K (10 μ g/100 ml in 20 mM Tris-HCl, 2 mM CaCl_2) at 37 degree Celsius for 8 min and dehydrated. The hybridization mixture containing the probe is placed on the slide, covered with a coverslip, sealed with rubber cement and incubated overnight in a humid chamber at 37 degree Celsius. After hybridization and post-hybridization washes, the biotinylated probe is detected by avidin-FITC and amplified with
35 additional layers of biotinylated goat anti-avidin and avidin-FITC. For chromosomal localization, fluorescent R-bands are obtained as previously described (Cherif *et al.*, 1990). The slides are

observed under a LEICA fluorescence microscope (DMRXA). Chromosomes are counterstained with propidium iodide and the fluorescent signal of the probe appears as two symmetrical yellow-green spots on both chromatids of the fluorescent R-band chromosome (red). Thus, a particular GENSET cDNA or genomic DNA may be localized to a particular cytogenetic R-band on a given
5 chromosome.

Use of cDNAs to Construct or Expand Chromosome Maps

Once the GENSET cDNAs or genomic DNAs have been assigned to particular chromosomes using any technique known to those skilled in the art those skilled in the art, particularly those described herein, they may be utilized to construct a high resolution map of the
10 chromosomes on which they are located or to identify the chromosomes in a sample.

Chromosome mapping involves assigning a given unique sequence to a particular chromosome as described above. Once the unique sequence has been mapped to a given chromosome, it is ordered relative to other unique sequences located on the same chromosome. One approach to chromosome mapping utilizes a series of yeast artificial chromosomes (YACs)
15 bearing several thousand long inserts derived from the chromosomes of the organism from which the GENSET cDNAs or genomic DNAs are obtained. This approach is described in Nagaraja *et al.* (1997), which disclosure is hereby incorporated by reference in its entirety. Briefly, in this approach each chromosome is broken into overlapping pieces which are inserted into the YAC vector. The YAC inserts are screened using PCR or other methods to determine whether they
20 include the GENSET cDNA or genomic DNA whose position is to be determined. Once an insert has been found which includes the GENSET cDNA or genomic DNA, the insert can be analyzed by PCR or other methods to determine whether the insert also contains other sequences known to be on the chromosome or in the region from which the GENSET cDNA or genomic DNA was derived. This process can be repeated for each insert in the YAC library to determine the location of each of
25 the GENSET cDNA or genomic DNA relative to one another and to other known chromosomal markers. In this way, a high resolution map of the distribution of numerous unique markers along each of the organisms chromosomes may be obtained.

Identification of genes associated with hereditary diseases or drug response

This example illustrates an approach useful for the association of GENSET cDNAs or
30 genomic DNAs with particular phenotypic characteristics. In this example, a particular GENSET cDNA or genomic DNA is used as a test probe to associate that GENSET cDNA or genomic DNA with a particular phenotypic characteristic.

GENSET cDNAs or genomic DNAs are mapped to a particular location on a human chromosome using techniques such as those described herein or other techniques known in the art.
35 A search of Mendelian Inheritance in Man (V. McKusick, Mendelian Inheritance in Man (available on line through Johns Hopkins University Welch Medical Library) reveals the region of the human

chromosome which contains the GENSET cDNA or genomic DNA to be a very gene rich region containing several known genes and several diseases or phenotypes for which genes have not been identified. The gene corresponding to this GENSET cDNA or genomic DNA thus becomes an immediate candidate for each of these genetic diseases.

- 5 Cells from patients with these diseases or phenotypes are isolated and expanded in culture. PCR primers from the GENSET cDNA or genomic DNA are used to screen genomic DNA, mRNA or cDNA obtained from the patients. GENSET cDNAs or genomic DNAs that are not amplified in the patients can be positively associated with a particular disease by further analysis. Alternatively, the PCR analysis may yield fragments of different lengths when the samples are derived from an
- 10 individual having the phenotype associated with the disease than when the sample is derived from a healthy individual, indicating that the gene containing the cDNA may be responsible for the genetic disease.

Uses of polynucleotides in recombinant vectors

- The present invention also relates to recombinant vectors, which include the isolated
- 15 polynucleotides of the present invention, or fragments thereof and to host cells recombinant for a polynucleotide of the invention, such as the above vectors, as well as to methods of making such vectors and host cells and for using them for production of GENSET polypeptides by recombinant techniques.

Recombinant Vectors

- 20 The term "vector" is used herein to designate either a circular or a linear DNA or RNA molecule, which is either double-stranded or single-stranded, and which comprise at least one polynucleotide of interest that is sought to be transferred in a cell host or in a unicellular or multicellular host organism. The present invention encompasses a family of recombinant vectors that comprise a regulatory polynucleotide and/or a coding polynucleotide derived from either the
- 25 GENSET genomic sequence or the cDNA sequence. Generally, a recombinant vector of the invention may comprise any of the polynucleotides described herein, including regulatory sequences, coding sequences and polynucleotide constructs, as well as any GENSET primer or probe as defined herein.

- In a first preferred embodiment, a recombinant vector of the invention is used to amplify the
- 30 inserted polynucleotide derived from a GENSET genomic sequence or a GENSET cDNA, for example any cDNA selected from the group consisting of sequences of SEQ ID Nos: 1-241, sequences of clone inserts of the deposited clone pool, variants and fragments thereof in a suitable cell host, this polynucleotide being amplified at every time that the recombinant vector replicates.

- A second preferred embodiment of the recombinant vectors according to the invention
- 35 comprises expression vectors comprising either a regulatory polynucleotide or a coding nucleic acid of the invention, or both. Within certain embodiments, expression vectors are employed to express

a GENSET polypeptide which can be then purified and, for example be used in ligand screening assays or as an immunogen in order to raise specific antibodies directed against the GENSET protein. In other embodiments, the expression vectors are used for constructing transgenic animals and also for gene therapy. Expression requires that appropriate signals are provided in the vectors, said signals including various regulatory elements, such as enhancers/promoters from both viral and mammalian sources that drive expression of the genes of interest in host cells. Dominant drug selection markers for establishing permanent, stable cell clones expressing the products are generally included in the expression vectors of the invention, as they are elements that link expression of the drug selection markers to expression of the polypeptide.

10 More particularly, the present invention relates to expression vectors which include nucleic acids encoding a GENSET protein, preferably a GENSET protein with an amino acid sequence selected from the group consisting of sequences of SEQ ID Nos: 242-482, mature polypeptides included in sequences of SEQ ID Nos: 242-272 and 274-384, and sequences of full-length or mature polypeptides encoded by the clone inserts of the deposited clone pool, as well as variants
15 and fragments thereof. The polynucleotides of the present invention may be used to express an encoded protein in a host organism to produce a beneficial effect. In such procedures, the encoded protein may be transiently expressed in the host organism or stably expressed in the host organism. The encoded protein may have any of the activities described herein. The encoded protein may be a protein which the host organism lacks or, alternatively, the encoded protein may augment the
20 existing levels of the protein in the host organism.

Some of the elements which can be found in the vectors of the present invention are described in further detail in the following sections.

General features of the expression vectors of the invention

A recombinant vector according to the invention comprises, but is not limited to, a YAC
25 (Yeast Artificial Chromosome), a BAC (Bacterial Artificial Chromosome), a phage, a phagemid, a cosmid, a plasmid or even a linear DNA molecule which may comprise a chromosomal, non-chromosomal, semi-synthetic and synthetic DNA. Such a recombinant vector can comprise a transcriptional unit comprising an assembly of:

(1) a genetic element or elements having a regulatory role in gene expression, for example
30 promoters or enhancers. Enhancers are cis-acting elements of DNA, usually from about 10 to 300 bp in length that act on the promoter to increase the transcription.

(2) a structural or coding sequence which is transcribed into mRNA and eventually translated into a polypeptide, said structural or coding sequence being operably linked to the regulatory elements described in (1); and

35 (3) appropriate transcription initiation and termination sequences. Structural units intended for use in yeast or eukaryotic expression systems preferably include a leader sequence enabling

extracellular secretion of translated protein by a host cell. Alternatively, when a recombinant protein is expressed without a leader or transport sequence, it may include a N-terminal residue. This residue may or may not be subsequently cleaved from the expressed recombinant protein to provide a final product.

5 Generally, recombinant expression vectors will include origins of replication, selectable markers permitting transformation of the host cell, and a promoter derived from a highly expressed gene to direct transcription of a downstream structural sequence. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably a leader sequence capable of directing secretion of the translated protein into the
10 periplasmic space or the extracellular medium. In a specific embodiment wherein the vector is adapted for transfecting and expressing desired sequences in mammalian host cells, preferred vectors will comprise an origin of replication in the desired host, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation signal, splice donor and acceptor sites, transcriptional termination sequences, and 5'-flanking non-transcribed sequences. DNA
15 sequences derived from the SV40 viral genome, for example SV40 origin, early promoter, enhancer, splice and polyadenylation signals may be used to provide the required non-transcribed genetic elements.

The *in vivo* expression of a GENSET polypeptide of the present invention may be useful in order to correct a genetic defect related to the expression of the native gene in a host organism or to
20 the production of a biologically inactive GENSET protein. Consequently, the present invention also comprises recombinant expression vectors mainly designed for the *in vivo* production of a GENSET polypeptide of the present invention by the introduction of the appropriate genetic material in the organism or the patient to be treated. This genetic material may be introduced *in vitro* in a cell that has been previously extracted from the organism, the modified cell being subsequently reintroduced
25 in the said organism, directly *in vivo* into the appropriate tissue.

Regulatory Elements

The suitable promoter regions used in the expression vectors according to the present invention are chosen taking into account the cell host in which the heterologous gene has to be expressed. The particular promoter employed to control the expression of a nucleic acid sequence
30 of interest is not believed to be important, so long as it is capable of directing the expression of the nucleic acid in the targeted cell. Thus, where a human cell is targeted, it is preferable to position the nucleic acid coding region adjacent to and under the control of a promoter that is capable of being expressed in a human cell, such as, for example, a human or a viral promoter.

A suitable promoter may be heterologous with respect to the nucleic acid for which it
35 controls the expression or alternatively can be endogenous to the native polynucleotide containing the coding sequence to be expressed. Additionally, the promoter is generally heterologous with

respect to the recombinant vector sequences within which the construct promoter/coding sequence has been inserted.

Promoter regions can be selected from any desired gene using, for example, CAT (chloramphenicol transferase) vectors and more preferably pKK232-8 and pCM7 vectors.

- 5 Preferred bacterial promoters are the LacI, LacZ, the T3 or T7 bacteriophage RNA polymerase promoters, the gpt, lambda PR, PL and trp promoters (EP 0036776), the polyhedrin promoter, or the p10 protein promoter from baculovirus (Kit Novagen), (Smith *et al.*, 1983; O'Reilly *et al.*, 1992), which disclosures are hereby incorporated by reference in their entireties, the lambda PR promoter or also the trc promoter.
- 10 Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-L. Selection of a convenient vector and promoter is well within the level of ordinary skill in the art. The choice of a promoter is well within the ability of a person skilled in the field of genetic engineering. For example, one may refer to the book of Sambrook *et al.*, (1989) or also to the procedures described by Fuller *et al.*, (1996), which
- 15 disclosures are hereby incorporated by reference in their entireties.

Other regulatory elements

- Where a cDNA insert is employed, one will typically desire to include a polyadenylation signal to effect proper polyadenylation of the gene transcript. The nature of the polyadenylation signal is not believed to be crucial to the successful practice of the invention, and any such sequence
- 20 may be employed such as human growth hormone and SV40 polyadenylation signals. Also contemplated as an element of the expression cassette is a terminator. These elements can serve to enhance message levels and to minimize read through from the cassette into other sequences.

Selectable Markers

- Selectable markers confer an identifiable change to the cell permitting easy identification of
- 25 cells containing the expression construct. The selectable marker genes for selection of transformed host cells are preferably dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, TRP1 for *S. cerevisiae* or tetracycline, rifampicin or ampicillin resistance in *E. Coli*, or levan saccharase for mycobacteria, this latter marker being a negative selection marker.

Preferred Vectors

30 Bacterial vectors

As a representative but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and a bacterial origin of replication derived from commercially available plasmids comprising genetic elements of pBR322 (ATCC 37017). Such commercial

vectors include, for example, pKK223-3 (Pharmacia, Uppsala, Sweden), and pGEM1 (Promega Biotec, Madison, WI, USA).

Large numbers of other suitable vectors are known to those of skill in the art, and commercially available, such as the following bacterial vectors: pQE70, pQE60, pQE-9 (Qiagen),
5 pbs, pD10, phagescript, psiX174, pbluescript SK, pbsks, pNH8A, pNH16A, pNH18A, pNH46A (Stratagene); ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia); pWLNEO, pSV2CAT, pOG44, pXT1, pSG (Stratagene); pSVK3, pBPV, pMSG, pSVL (Pharmacia); pQE-30 (QIAexpress).

Bacteriophage vectors

10 The P1 bacteriophage vector may contain large inserts ranging from about 80 to about 100 kb. The construction of P1 bacteriophage vectors such as p158 or p158/neo8 are notably described by Sternberg (1992, 1994), which disclosure is hereby incorporated by reference in its entirety. Recombinant P1 clones comprising GENSET nucleotide sequences may be designed for inserting large polynucleotides of more than 40 kb (See Linton *et al.*, 1993), which disclosure is hereby
15 incorporated by reference in its entirety. To generate P1 DNA for transgenic experiments, a preferred protocol is the protocol described by McCormick *et al.* (1994), which disclosure is hereby incorporated by reference in its entirety. Briefly, *E. coli* (preferably strain NS3529) harboring the P1 plasmid are grown overnight in a suitable broth medium containing 25 µg/ml of kanamycin. The P1 DNA is prepared from the *E. coli* by alkaline lysis using the Qiagen Plasmid Maxi kit
20 (Qiagen, Chatsworth, CA, USA), according to the manufacturer's instructions. The P1 DNA is purified from the bacterial lysate on two Qiagen-tip 500 columns, using the washing and elution buffers contained in the kit. A phenol/chloroform extraction is then performed before precipitating the DNA with 70% ethanol. After solubilizing the DNA in TE (10 mM Tris-HCl, pH 7.4, 1 mM EDTA), the concentration of the DNA is assessed by spectrophotometry.

25 When the goal is to express a P1 clone comprising GENSET nucleotide sequences in a transgenic animal, typically in transgenic mice, it is desirable to remove vector sequences from the P1 DNA fragment, for example by cleaving the P1 DNA at rare-cutting sites within the P1 polylinker (*SfiI*, *NotI* or *SalI*). The P1 insert is then purified from vector sequences on a pulsed-field agarose gel, using methods similar to those originally reported for the isolation of DNA from
30 YACs (See *e. g.*, Schedl *et al.*, 1993a; Peterson *et al.*, 1993), which disclosures are hereby incorporated by reference in their entireties. At this stage, the resulting purified insert DNA can be concentrated, if necessary, on a Millipore Ultrafree-MC Filter Unit (Millipore, Bedford, MA, USA – 30,000 molecular weight limit) and then dialyzed against microinjection buffer (10 mM Tris-HCl, pH 7.4; 250 µM EDTA) containing 100 mM NaCl, 30 µM spermine, 70 µM spermidine on a
35 microdialysis membrane (type VS, 0.025 µM from Millipore). The intactness of the purified P1

DNA insert is assessed by electrophoresis on 1% agarose (Sea Kem GTG; FMC Bio-products) pulse-field gel and staining with ethidium bromide.

Viral vectors

In one specific embodiment, the vector is derived from an adenovirus. Preferred adenovirus
5 vectors according to the invention are those described by Feldman and Steg (1996), or Ohno *et al.*,
(1994), which disclosures are hereby incorporated by reference in their entireties. Another
preferred recombinant adenovirus according to this specific embodiment of the present invention is
the human adenovirus type 2 or 5 (Ad 2 or Ad 5) or an adenovirus of animal origin (French patent
application No. FR-93.05954), which disclosure is hereby incorporated by reference in its entirety.

10 Retrovirus vectors and adeno-associated virus vectors are generally understood to be the
recombinant gene delivery systems of choice for the transfer of exogenous polynucleotides *in vivo*,
particularly to mammals, including humans. These vectors provide efficient delivery of genes into
cells, and the transferred nucleic acids are stably integrated into the chromosomal DNA of the host.
Particularly preferred retroviruses for the preparation or construction of retroviral *in vitro* or *in vitro*
15 gene delivery vehicles of the present invention include retroviruses selected from the group
consisting of Mink-Cell Focus Inducing Virus, Murine Sarcoma Virus, Reticuloendotheliosis virus
and Rous Sarcoma virus. Particularly preferred Murine Leukemia Viruses include the 4070A and
the 1504A viruses, Abelson (ATCC No VR-999), Friend (ATCC No VR-245), Gross (ATCC No
VR-590), Rauscher (ATCC No VR-998) and Moloney Murine Leukemia Virus (ATCC No VR-
20 190; PCT Application No WO 94/24298). Particularly preferred Rous Sarcoma Viruses include
Bryan high titer (ATCC Nos VR-334, VR-657, VR-726, VR-659 and VR-728). Other preferred
retroviral vectors are those described in Roth *et al.* (1996), PCT Application No WO 93/25234,
PCT Application No WO 94/06920, Roux *et al.*, (1989), Julan *et al.*, (1992), and Neda *et al.*,
(1991), which disclosures are hereby incorporated by reference in their entireties.

25 Yet another viral vector system that is contemplated by the invention comprises the adeno-
associated virus (AAV). The adeno-associated virus is a naturally occurring defective virus that
requires another virus, such as an adenovirus or a herpes virus, as a helper virus for efficient
replication and a productive life cycle (Muzyczka *et al.*, 1992), which disclosure is hereby
incorporated by reference in its entirety. It is also one of the few viruses that may integrate its DNA
30 into non-dividing cells, and exhibits a high frequency of stable integration (Flotte *et al.* 1992;
Samulski *et al.*, 1989; McLaughlin *et al.*, 1989), which disclosures are hereby incorporated by
reference in their entireties. One advantageous feature of AAV derives from its reduced efficacy
for transducing primary cells relative to transformed cells.

BAC vectors

35 The bacterial artificial chromosome (BAC) cloning system (Shizuya *et al.*, 1992), which
disclosure is hereby incorporated by reference in its entirety, has been developed to stably maintain

large fragments of genomic DNA (100-300 kb) in *E. coli*. A preferred BAC vector comprises a pBeloBAC11 vector that has been described by Kim *et al.* (1996), which disclosure is hereby incorporated by reference in its entirety. BAC libraries are prepared with this vector using size-selected genomic DNA that has been partially digested using enzymes that permit ligation into
5 either the *Bam* HI or *Hind*III sites in the vector. Flanking these cloning sites are T7 and SP6 RNA polymerase transcription initiation sites that can be used to generate end probes by either RNA transcription or PCR methods. After the construction of a BAC library in *E. coli*, BAC DNA is purified from the host cell as a supercoiled circle. Converting these circular molecules into a linear form precedes both size determination and introduction of the BACs into recipient cells. The
10 cloning site is flanked by two *Not* I sites, permitting cloned segments to be excised from the vector by *Not* I digestion. Alternatively, the DNA insert contained in the pBeloBAC11 vector may be linearized by treatment of the BAC vector with the commercially available enzyme lambda terminase that leads to the cleavage at the unique *cos*N site, but this cleavage method results in a full length BAC clone containing both the insert DNA and the BAC sequences.

15 Baculovirus:

Another specific suitable host vector system is the pVL1392/1393 baculovirus transfer vector (Pharminogen) that is used to transfect the SF9 cell line (ATCC No. CRL 1711) which is derived from *Spodoptera frugiperda*. Other suitable vectors for the expression of the GENSET polypeptide of the present invention in a baculovirus expression system include those described by
20 Chai *et al.*, (1993), Vlasak *et al.*, (1983), and Lenhard *et al.*, (1996), which disclosures are hereby incorporated by reference in their entireties.

Delivery Of The Recombinant Vectors:

To effect expression of the polynucleotides and polynucleotide constructs of the invention, these constructs must be delivered into a cell. This delivery may be accomplished *in vitro*, as in
25 laboratory procedures for transforming cell lines, or *in vivo* or *ex vivo*, as in the treatment of certain diseases states. One mechanism is viral infection where the expression construct is encapsulated in an infectious viral particle.

Several non-viral methods for the transfer of polynucleotides into cultured mammalian cells are also contemplated by the present invention, and include, without being limited to, calcium
30 phosphate precipitation (Graham *et al.*, 1973; Chen *et al.*, 1987); DEAE-dextran (Gopal, 1985); electroporation (Tur-Kaspa *et al.*, 1986; Potter *et al.*, 1984); direct microinjection (Harland *et al.*, 1985); DNA-loaded liposomes (Nicolau *et al.*, 1982; Fraley *et al.*, 1979); and receptor-mediated transfection. (Wu and Wu, 1987, 1988), which disclosures are hereby incorporated by reference in their entireties. Some of these techniques may be successfully adapted for *in vivo* or *ex vivo* use.

35 Once the expression polynucleotide has been delivered into the cell, it may be stably integrated into the genome of the recipient cell. This integration may be in the cognate location and

orientation via homologous recombination (gene replacement) or it may be integrated in a random, non-specific location (gene augmentation). In yet further embodiments, the nucleic acid may be stably maintained in the cell as a separate, episomal segment of DNA. Such nucleic acid segments or "episomes" encode sequences sufficient to permit maintenance and replication independent of or
5 in synchronization with the host cell cycle.

One specific embodiment for a method for delivering a protein or peptide to the interior of a cell of a vertebrate *in vivo* comprises the step of introducing a preparation comprising a physiologically acceptable carrier and a naked polynucleotide operatively coding for the polypeptide of interest into the interstitial space of a tissue comprising the cell, whereby the naked
10 polynucleotide is taken up into the interior of the cell and has a physiological effect. This is particularly applicable for transfer *in vitro* but it may be applied to *in vivo* as well.

Compositions for use *in vitro* and *in vivo* comprising a "naked" polynucleotide are described in PCT application No. WO 90/11092 (Vical Inc.) and also in PCT application No. WO 95/11307 (Institut Pasteur, INSERM, Université d'Ottawa) as well as in the articles of Tacson *et al.*
15 (1996) and of Huygen *et al.*, (1996), which disclosures are hereby incorporated by reference in their entireties.

In still another embodiment of the invention, the transfer of a naked polynucleotide of the invention, including a polynucleotide construct of the invention, into cells may be proceeded with a particle bombardment (biolistic), said particles being DNA-coated microprojectiles accelerated to a
20 high velocity allowing them to pierce cell membranes and enter cells without killing them, such as described by Klein *et al.*, (1987), which disclosure is hereby incorporated by reference in its entirety.

In a further embodiment, the polynucleotide of the invention may be entrapped in a liposome (Ghosh and Bacchawat, 1991; Wong *et al.*, 1980; Nicolau *et al.*, 1987), which disclosures
25 are hereby incorporated by reference in their entireties.

In a specific embodiment, the invention provides a composition for the *in vivo* production of the GENSET protein or polypeptide described herein. It comprises a naked polynucleotide operatively coding for this polypeptide, in solution in a physiologically acceptable carrier, and suitable for introduction into a tissue to cause cells of the tissue to express the said protein or
30 polypeptide.

The amount of vector to be injected to the desired host organism varies according to the site of injection. As an indicative dose, it will be injected between 0,1 and 100 µg of the vector in an animal body, preferably a mammal body, for example a mouse body.

In another embodiment of the vector according to the invention, it may be introduced *in*
35 *vitro* in a host cell, preferably in a host cell previously harvested from the animal to be treated and more preferably a somatic cell such as a muscle cell. In a subsequent step, the cell that has been transformed with the vector coding for the desired GENSET polypeptide or the desired fragment

thereof is reintroduced into the animal body in order to deliver the recombinant protein within the body either locally or systemically.

Secretion vectors

Some of the GENSET cDNAs or genomic DNAs of the invention may also be used to
5 construct secretion vectors capable of directing the secretion of the proteins encoded by genes inserted in the vectors. Such secretion vectors may facilitate the purification or enrichment of the proteins encoded by genes inserted therein by reducing the number of background proteins from which the desired protein must be purified or enriched. Exemplary secretion vectors are described below.

10 The secretion vectors of the present invention include a promoter capable of directing gene expression in the host cell, tissue, or organism of interest. Such promoters include the Rous Sarcoma Virus promoter, the SV40 promoter, the human cytomegalovirus promoter, and other promoters familiar to those skilled in the art.

A signal sequence from a polynucleotide of the invention, preferably a signal sequences
15 selected from the group of signal sequences of SEQ ID Nos: 1-31 and 33-143 and signal sequences of clone inserts of the deposited clone pool is operably linked to the promoter such that the mRNA transcribed from the promoter will direct the translation of the signal peptide. The host cell, tissue, or organism may be any cell, tissue, or organism which recognizes the signal peptide encoded by the signal sequence in the GENSET cDNA or genomic DNA. Suitable hosts include mammalian
20 cells, tissues or organisms, avian cells, tissues, or organisms, insect cells, tissues or organisms, or yeast.

In addition, the secretion vector contains cloning sites for inserting genes encoding the proteins which are to be secreted. The cloning sites facilitate the cloning of the insert gene in frame with the signal sequence such that a fusion protein in which the signal peptide is fused to the protein
25 encoded by the inserted gene is expressed from the mRNA transcribed from the promoter. The signal peptide directs the extracellular secretion of the fusion protein.

The secretion vector may be DNA or RNA and may integrate into the chromosome of the host, be stably maintained as an extrachromosomal replicon in the host, be an artificial chromosome, or be transiently present in the host. Preferably, the secretion vector is maintained in
30 multiple copies in each host cell. As used herein, multiple copies means at least 2, 5, 10, 20, 25, 50 or more than 50 copies per cell. In some embodiments, the multiple copies are maintained extrachromosomally. In other embodiments, the multiple copies result from amplification of a chromosomal sequence.

Many nucleic acid backbones suitable for use as secretion vectors are known to those
35 skilled in the art, including retroviral vectors, SV40 vectors, Bovine Papilloma Virus vectors, yeast integrating plasmids, yeast episomal plasmids, yeast artificial chromosomes, human artificial

chromosomes, P element vectors, baculovirus vectors, or bacterial plasmids capable of being transiently introduced into the host.

The secretion vector may also contain a polyA signal such that the polyA signal is located downstream of the gene inserted into the secretion vector.

- 5 After the gene encoding the protein for which secretion is desired is inserted into the secretion vector, the secretion vector is introduced into the host cell, tissue, or organism using calcium phosphate precipitation, DEAE-Dextran, electroporation, liposome-mediated transfection, viral particles or as naked DNA. The protein encoded by the inserted gene is then purified or enriched from the supernatant using conventional techniques such as ammonium sulfate
- 10 precipitation, immunoprecipitation, immunochromatography, size exclusion chromatography, ion exchange chromatography, and hplc. Alternatively, the secreted protein may be in a sufficiently enriched or pure state in the supernatant or growth media of the host to permit it to be used for its intended purpose without further enrichment.

- The signal sequences may also be inserted into vectors designed for gene therapy. In such
- 15 vectors, the signal sequence is operably linked to a promoter such that mRNA transcribed from the promoter encodes the signal peptide. A cloning site is located downstream of the signal sequence such that a gene encoding a protein whose secretion is desired may readily be inserted into the vector and fused to the signal sequence. The vector is introduced into an appropriate host cell. The protein expressed from the promoter is secreted extracellularly, thereby producing a therapeutic
- 20 effect.

Cell Hosts

- Another object of the invention comprises a host cell that has been transformed or transfected with one of the polynucleotides described herein, and in particular a polynucleotide either comprising a GENSET regulatory polynucleotide or the polynucleotide coding for a
- 25 GENSET polypeptide. Also included are host cells that are transformed (prokaryotic cells) or that are transfected (eukaryotic cells) with a recombinant vector such as one of those described above. However, the cell hosts of the present invention can comprise any of the polynucleotides of the present invention. In a preferred embodiment, host cells contain a polynucleotide sequence comprising a sequence selected from the group consisting of sequences of SEQ ID Nos: 1-241,
- 30 sequences of clone inserts of the deposited clone pool, variants and fragments thereof. Preferred host cells used as recipients for the expression vectors of the invention are the following:

- a) Prokaryotic host cells: *Escherichia coli* strains (I.E.DH5- α strain), *Bacillus subtilis*, *Salmonella typhimurium*, and strains from species like *Pseudomonas*, *Streptomyces* and *Staphylococcus*.
- 35 b) Eukaryotic host cells: HeLa cells (ATCC No.CCL2; No.CCL2.1; No.CCL2.2), Cv 1 cells (ATCC No.CCL70), COS cells (ATCC No.CRL1650; No.CRL1651), Sf-9 cells (ATCC

No.CRL1711), C127 cells (ATCC No. CRL-1804), 3T3 (ATCC No. CRL-6361), CHO (ATCC No. CCL-61), human kidney 293. (ATCC No. 45504; No. CRL-1573) and BHK (ECACC No. 84100501; No. 84111301).

c) Other mammalian host cells.

- 5 The present invention also encompasses primary, secondary, and immortalized homologously recombinant host cells of vertebrate origin, preferably mammalian origin and particularly human origin, that have been engineered to: a) insert exogenous (heterologous) polynucleotides into the endogenous chromosomal DNA of a targeted gene, b) delete endogenous chromosomal DNA, and/or c) replace endogenous chromosomal DNA with exogenous
- 10 polynucleotides. Insertions, deletions, and/or replacements of polynucleotide sequences may be to the coding sequences of the targeted gene and/or to regulatory regions, such as promoter and enhancer sequences, operably associated with the targeted gene.

- In addition to encompassing host cells containing the vector constructs discussed herein, the invention also encompasses primary, secondary, and immortalized host cells of vertebrate origin,
- 15 particularly mammalian origin, that have been engineered to delete or replace endogenous genetic material (e.g., coding sequence), and/or to include genetic material (e.g., heterologous polynucleotide sequences) that is operably associated with the polynucleotides of the invention, and which activates, alters, and/or amplifies endogenous polynucleotides. For example, techniques known in the art may be used to operably associate heterologous control regions (e.g., promoter
- 20 and/or enhancer) and endogenous polynucleotide sequences via homologous recombination, see, e.g., U.S. Patent No. 5,641,670, issued June 24, 1997; International Publication No. WO 96/29411, published September 26, 1996; International Publication No. WO 94/12650, published August 4, 1994; Koller *et al.*, (1989); and Zijlstra *et al.* (1989) (The disclosures of each of which are incorporated by reference in their entireties).

- 25 The present invention further relates to a method of making a homologously recombinant host cell *in vitro* or *in vivo*, wherein the expression of a targeted gene not normally expressed in the cell is altered. Preferably the alteration causes expression of the targeted gene under normal growth conditions or under conditions suitable for producing the polypeptide encoded by the targeted gene. The method comprises the steps of: (a) transfecting the cell *in vitro* or *in vivo* with a polynucleotide
- 30 construct, said polynucleotide construct comprising: (i) a targeting sequence; (ii) a regulatory sequence and/or a coding sequence; and (iii) an unpaired splice donor site, if necessary, thereby producing a transfected cell; and (b) maintaining the transfected cell *in vitro* or *in vivo* under conditions appropriate for homologous recombination.

- The present invention further relates to a method of altering the expression of a targeted
- 35 gene in a cell *in vitro* or *in vivo* wherein the gene is not normally expressed in the cell, comprising the steps of: (a) transfecting the cell *in vitro* or *in vivo* with a polynucleotide construct, said polynucleotide construct comprising: (i) a targeting sequence; (ii) a regulatory sequence and/or a

coding sequence; and (iii) an unpaired splice donor site, if necessary, thereby producing a transfected cell; and (b) maintaining the transfected cell *in vitro* or *in vivo* under conditions appropriate for homologous recombination, thereby producing a homologously recombinant cell; and (c) maintaining the homologously recombinant cell *in vitro* or *in vivo* under conditions

5 appropriate for expression of the gene.

The present invention further relates to a method of making a polypeptide of the present invention by altering the expression of a targeted endogenous gene in a cell *in vitro* or *in vivo* wherein the gene is not normally expressed in the cell, comprising the steps of: a) transfecting the cell *in vitro* with a polynucleotide construct, said polynucleotide construct comprising: (i) a
10 targeting sequence; (ii) a regulatory sequence and/or a coding sequence; and (iii) an unpaired splice donor site, if necessary, thereby producing a transfected cell; (b) maintaining the transfected cell *in vitro* or *in vivo* under conditions appropriate for homologous recombination, thereby producing a homologously recombinant cell; and c) maintaining the homologously recombinant cell *in vitro* or *in vivo* under conditions appropriate for expression of the gene thereby making the polypeptide.

15 The present invention further relates to a polynucleotide construct which alters the expression of a targeted gene in a cell type in which the gene is not normally expressed. This occurs when the polynucleotide construct is inserted into the chromosomal DNA of the target cell, wherein said polynucleotide construct comprises: a) a targeting sequence; b) a regulatory sequence and/or coding sequence; and c) an unpaired splice-donor site, if necessary. Further included are a
20 polynucleotide construct, as described above, wherein said polynucleotide construct further comprises a polynucleotide which encodes a polypeptide and is in-frame with the targeted endogenous gene after homologous recombination with chromosomal DNA.

The compositions may be produced, and methods performed, by techniques known in the art, such as those described in U.S. Patent Nos: 6,054,288; 6,048,729; 6,048,724; 6,048,524;
25 5,994,127; 5,968,502; 5,965,125; 5,869,239; 5,817,789; 5,783,385; 5,733,761; 5,641,670; 5,580,734 ; International Publication Nos: WO96/29411, WO 94/12650; and scientific articles described by Koller *et al.*, (1994). (The disclosures of each of which are incorporated by reference in their entireties).

The GENSET gene expression in mammalian cells, preferably human cells, may be
30 rendered defective, or alternatively may be altered by replacing the endogenous GENSET gene in the genome of an animal cell by a GENSET polynucleotide according to the invention. These genetic alterations may be generated by homologous recombination using previously described specific polynucleotide constructs.

Mammal zygotes, such as murine zygotes may be used as cell hosts. For example, murine
35 zygotes may undergo microinjection with a purified DNA molecule of interest, for example a purified DNA molecule that has previously been adjusted to a concentration ranging from 1 ng/ml – for BAC inserts- to 3 ng/μl –for P1 bacteriophage inserts- in 10 mM Tris-HCl, pH 7.4, 250 μM

EDTA containing 100 mM NaCl, 30 μ M spermine, and 70 μ M spermidine. When the DNA to be microinjected has a large size, polyamines and high salt concentrations can be used in order to avoid mechanical breakage of this DNA, as described by Schedl *et al* (1993b), which disclosure is hereby incorporated by reference in its entirety.

- 5 Any one of the polynucleotides of the invention, including the Polynucleotide constructs described herein, may be introduced in an embryonic stem (ES) cell line, preferably a mouse ES cell line. ES cell lines are derived from pluripotent, uncommitted cells of the inner cell mass of pre-implantation blastocysts. Preferred ES cell lines are the following: ES-E14TG2a (ATCC No.CRL-1821), ES-D3 (ATCC No.CRL1934 and No. CRL-11632), YS001 (ATCC No. CRL-11776), 36.5
- 10 (ATCC No. CRL-11116). ES cells are maintained in an uncommitted state by culture in the presence of growth-inhibited feeder cells which provide the appropriate signals to preserve this embryonic phenotype and serve as a matrix for ES cell adherence. Preferred feeder cells are primary embryonic fibroblasts that are established from tissue of day 13- day 14 embryos of virtually any mouse strain, that are maintained in culture, such as described by Abbondanzo *et al.*
- 15 (1993) and are growth-inhibited by irradiation, such as described by Robertson (1987), or by the presence of an inhibitory concentration of LIF, such as described by Pease and Williams (1990), which disclosures are hereby incorporated by reference in their entireties.

The constructs in the host cells can be used in a conventional manner to produce the gene product encoded by the recombinant sequence.

- 20 Following transformation of a suitable host and growth of the host to an appropriate cell density, the selected promoter is induced by appropriate means, such as temperature shift or chemical induction, and cells are cultivated for an additional period. Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification. Microbial cells employed in the expression of proteins can be disrupted by
- 25 any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents. Such methods are well known by the skilled artisan.

Transgenic Animals

- The terms "transgenic animals" or "host animals" are used herein to designate animals that have their genome genetically and artificially manipulated so as to include one of the nucleic acids
- 30 according to the invention. Preferred animals are non-human mammals and include those belonging to a genus selected from *Mus* (e.g. mice), *Rattus* (e.g. rats) and *Oryctogalus* (e.g. rabbits) which have their genome artificially and genetically altered by the insertion of a nucleic acid according to the invention. In one embodiment, the invention encompasses non-human host mammals and animals comprising a recombinant vector of the invention or a GENSET gene
- 35 disrupted by homologous recombination with a knock out vector.

The transgenic animals of the invention all include within a plurality of their cells a cloned recombinant or synthetic DNA sequence, more specifically one of the purified or isolated nucleic acids comprising a GENSET coding sequence, a GENSET regulatory polynucleotide, a polynucleotide construct, or a DNA sequence encoding an antisense polynucleotide such as
5 described in the present specification.

Generally, a transgenic animal according the present invention comprises any of the polynucleotides, the recombinant vectors and the cell hosts described in the present invention. In a first preferred embodiment, these transgenic animals may be good experimental models in order to study the diverse pathologies related to the dysregulation of the expression of a given GENSET
10 gene, in particular the transgenic animals containing within their genome one or several copies of an inserted polynucleotide encoding a native GENSET protein, or alternatively a mutant GENSET protein.

In a second preferred embodiment, these transgenic animals may express a desired polypeptide of interest under the control of the regulatory polynucleotides of the GENSET gene,
15 leading to high yields in the synthesis of this protein of interest, and eventually to tissue specific expression of the protein of interest.

In a third preferred embodiment, these transgenic animals may express a desired polypeptide of interest fused to a GENSET signal peptide sequence, leading to the secretion of the fusion (chimeric) polypeptide.

20 The design of the transgenic animals of the invention may be made according to the conventional techniques well known from the one skilled in the art. For more details regarding the production of transgenic animals, and specifically transgenic mice, it may be referred to US Patents Nos 4,873,191, issued Oct. 10, 1989; 5,464,764 issued Nov 7, 1995; and 5,789,215, issued Aug 4, 1998; these documents being herein incorporated by reference to disclose methods producing
25 transgenic mice.

Transgenic animals of the present invention are produced by the application of procedures which result in an animal with a genome that has incorporated exogenous genetic material. The procedure involves obtaining the genetic material which encodes either a GENSET coding sequence, a GENSET regulatory polynucleotide or a DNA sequence encoding a GENSET antisense
30 polynucleotide, or a portion thereof, such as described in the present specification. A recombinant polynucleotide of the invention is inserted into an embryonic or ES stem cell line. The insertion is preferably made using electroporation, such as described by Thomas *et al.* (1987), which disclosure is hereby incorporated by reference in its entirety. The cells subjected to electroporation are screened (e.g. by selection via selectable markers, by PCR or by Southern blot analysis) to find
35 positive cells which have integrated the exogenous recombinant polynucleotide into their genome, preferably via an homologous recombination event. An illustrative positive-negative selection

procedure that may be used according to the invention is described by Mansour *et al.* (1988), which disclosure is hereby incorporated by reference in its entirety.

The positive cells are then isolated, cloned and injected into 3.5 days old blastocysts from mice, such as described by Bradley (1987), which disclosure is hereby incorporated by reference in its entirety. The blastocysts are then inserted into a female host animal and allowed to grow to term. Alternatively, the positive ES cells are brought into contact with embryos at the 2.5 days old 8-16 cell stage (morulae) such as described by Wood *et al.* (1993), or by Nagy *et al.* (1993), which disclosures are hereby incorporated by reference in their entireties, the ES cells being internalized to colonize extensively the blastocyst including the cells which will give rise to the germ line.

The offspring of the female host are tested to determine which animals are transgenic e.g. include the inserted exogenous DNA sequence and which ones are wild type.

Thus, the present invention also concerns a transgenic animal containing a nucleic acid, a recombinant expression vector or a recombinant host cell according to the invention.

Recombinant Cell Lines Derived From The Transgenic Animals Of The Invention:

A further object of the invention comprises recombinant host cells obtained from a transgenic animal described herein. In one embodiment the invention encompasses cells derived from non-human host mammals and animals comprising a recombinant vector of the invention or a GENSET gene disrupted by homologous recombination with a knock out vector.

Recombinant cell lines may be established *in vitro* from cells obtained from any tissue of a transgenic animal according to the invention, for example by transfection of primary cell cultures with vectors expressing *onc*-genes such as SV40 large T antigen, as described by Chou (1989), and Shay *et al.* (1991), which disclosures are hereby incorporated by reference in their entireties.

USES OF POLYPEPTIDES OF THE INVENTION

Proteins containing multimerization domains

The invention relates to compositions and methods using proteins of the invention containing a multimerization domains such as a leucine zipper or a helix loop helix domain.

Proteins of the invention containing a leucine zipper domain, are herein referred to as LZP, such as the ones described in this section and those containing a leucine zipper domain as shown on Table VI, or parts thereof, preferably fragments comprising a leucine zipper domain, or derivative thereof to mediate multimerization of proteins of interest.

The leucine zipper consists of a periodic repetition of leucine residues at every seventh, covering a distance spanning eight helical turns. The segments containing these periodic arrays of leucine residues appear to exist in an alpha-helical conformation, and the leucine side chains extending from one alpha-helix interact with those from a similar alpha helix of a second

polypeptide, facilitating dimerization. The structure formed by cooperation of these two regions forms a coiled coil (O'Shea E.K., Rutkowski R., Kim P.S. *Science* 243:538-542.,1989).

Leucine-zippers contribute to targeting of various proteins (eg. glucose transporters, Asano, et al., *J. Biol. Chem.*, 267, 19636-19641 (1992)) and permit dimerization of various cytoplasmic hormone receptors and enzymes (Forman, et al., *Mol Endocrinol*, 3, 1610-1626 (1989)). Leucine zippers are also a common feature of protein transcription factors, where they permit homo- or heterodimerization resulting in tight binding to DNA strands (for reviews, see Abel, et al., *Nature* 341, 24-25 (1989); Jones, et al., *Cell* 61, 9-11 (1990); Lamb, et al., *Trends in Biochemical Sciences* 16, 417-422 (1991)).

Leucine zippers have been shown to be useful tools in several areas of biotechnology, especially in protein engineering, where their ability to mediate homo-dimerization or hetero-dimerization has found several applications. For example, Bosslet et al have described the use of a pair of leucine zipper for in vitro diagnosis, in particular for the immunochemical detection and determination of an analyte in a biological liquid (US patent 5,643,731) / Tso et al have used leucine zippers for producing bispecific antibody heterodimers (US patent 5,932,448) / Methods of preparing soluble oligomeric proteins using leucine zippers have been described by Conrad et al (US patent 5,965,712), Ciardelli et al (US patent 5,837,816), Spriggs et al (WO9410308) / Leucine zipper forming sequences have been used by Pelletier et al in protein fragment complementation assays to detect biomolecular interactions (WO9834120). Because of their usefulness in biotechnology, it is thus highly interesting to isolate new leucine zipper domains.

The multimerization activity of proteins containing leucine zipper domains may be assayed using any of the assays known to those skilled in the art including circular dichroism spectrum and thermal melting analyses as described in US patent 5,942,433. Alternatively, the leucine zipper motif in LZP could be used by those skilled in art as a "bait protein" in a well established yeast double hybridization system to identify its interacting protein partners in vivo from cDNA library derived from different tissues or cell types of a given organism. Alternatively, LZP or part thereof could be used by those skilled in art in mammalian cell transfection experiments. When fused to a suitable peptide tag such as [His]₆ tag in a protein expression vector and introduced into culture cells, this expressed fusion protein can be immunoprecipitated with its potential interacting proteins by using anti-tag peptide antibody. This method could be chosen either to identify the associated partner or to confirm the results obtained by other methods such as those just mentioned.

In a preferred embodiment, the invention relates to compositions and methods of using LZP or part thereof for preparing soluble multimeric proteins, which consist in multimers of fusion proteins containing a leucine zipper fused to a protein of interest, using any technique known to those skilled in the art including those described in international patent WO9410308, which disclosure is hereby incorporated by reference in its entirety. In another preferred embodiment, LZP or derivative thereof is used to produce bispecific antibody heterodimers as described in US

patent 5,932,448, which disclosure is hereby incorporated by reference in its entirety. Briefly, leucine zippers capable of forming heterodimers are respectively linked to epitope binding components with different specificities. Bispecific antibodies are formed by pairwise association of the leucine zippers, forming an heterodimer which links two distinct epitope binding components.

- 5 In still another preferred embodiment, LZP or part thereof or derivative thereof is used for detection and determination of an analyte in a biological liquid as described in US patent 5,643,731, which disclosure is hereby incorporated by reference in its entirety. Briefly, a first leucine zipper is immobilized on a solid support and the second leucine zipper is coupled to a specific binding partner for an analyte in a biological fluid. The two peptides are then brought into contact thereby
- 10 immobilizing the binding partner on the solid phase. The biological sample is then contacted with the immobilized binding partner and the amount of analyte in the sample bound to the binding partner determined. In still another preferred embodiment, the LZP or part thereof may be used to synthesize novel nucleic acid binding proteins which are able to multimerize with proteins of interest, for example to inhibit and/or control cellular growth using any genetic engineering
- 15 technique known to those skilled in the art including the ones described in the US patent 5,942,433, which disclosure is hereby incorporated by reference in its entirety .

In another embodiment, the invention relates to compositions and methods using the LZP or part thereof or derivative thereof in protein fragment complementation assays to detect biomolecular interactions in vivo and in vitro as described in international patent WO9834120,

- 20 which disclosures is hereby incorporated by reference in its entirety. Such assays may be used to study the equilibrium and kinetic aspects of molecular interactions including protein-protein, protein-nucleic acid, protein-carbohydrate and protein-small molecule interactions, for screening cDNA libraries for binding to a target protein with unknown proteins or libraries of small organic molecules for biological activity.

- 25 Still, another object of the present invention relates to the use of the LZP or part thereof for identifying new leucine zipper domains using any techniques for detecting protein-protein interaction known to those skilled in the art. Among the traditional methods which may be employed are co-immunoprecipitation, crosslinking and co-purification through gradients or chromatographic columns of cell lysates. Once isolated as a protein interacting with the LZP, such
- 30 an intracellular protein can be identified (e.g. its amino acid sequence determined) and can, in turn, be used, in conjunction with standard techniques, to identify other proteins with which it interacts. The amino acid sequence thus obtained may be used as a guide for the generation of oligonucleotide mixtures that can be used to screen for gene sequences encoding such intracellular proteins. Screening may be accomplished, for example, by standard hybridization or PCR techniques.
- 35 Techniques for the generation of oligonucleotide mixtures and the screening are well-known. (See, e.g., Ausubel *et al.*, eds., *Current Protocols in Molecular Biology*, J.Wiley and Sons (New York,

NY 1993) and PR Protocols: A Guide to Methods and Applications, 1990, Innis, M. et al., eds. Academic Press, Inc., New York).

Alternatively, methods may be employed which result in the simultaneous identification of genes which encode the intracellular proteins that can dimerize with the LZIP or part thereof using any technique known to those skilled in the art. These methods include, for example, probing cDNA expression libraries, in a manner similar to the well known technique of antibody probing of lambda.gt11 libraries, using as a probe a labeled version of the LZIP or part thereof, or fusion protein, e.g., the LZIP or part thereof fused to a marker (e.g., an enzyme, fluor, luminescent protein, or dye), or an Ig-Fc domain (for technical details on screening of cDNA expression libraries, see Ausubel *et al, supra*). Alternatively, another method for the detection of protein interaction in vivo, the two-hybrid system, may be used.

Protein of SEQ ID NO:261 (internal designation 116-054-3-0-E6-CS)

The 233 amino acids protein of SEQ ID NO: 261 encoded by the cDNA of SEQ ID NO: 20 displays two leucine zipper sites at positions 142-163 and 170-191.

It is believed that the protein of SEQ ID NO: 261 is able to dimerize either with itself (homo-dimerisation) or with an heterologous protein (hetero-dimerisation) of interest, through the mediation of its leucine zipper domain. Preferred polypeptides of the invention are polypeptides comprising fragments of SEQ ID NO: 261 from position 142-163 and 170-191, and fragments having any of the biological activities described herein.

Protein of SEQ ID NO:263 (internal designation 116-055-2-0-F7-CS)

The protein of SEQ ID NO: 263 encoded by the cDNA of SEQ ID NO: 22 displays a leucine zipper pattern situated near its NH2 terminal part (position 15 to 36).

It is believed that the protein of SEQ ID NO: 263 is able to dimerize either with itself (homo-dimerisation) or with an heterologous protein (hetero-dimerisation) of interest, through the mediation of its leucine zipper domain. Preferred polypeptides of the invention are polypeptides comprising fragments of SEQ ID NO: 263 from position 15 to 36, and fragments having any of the biological activities described herein..

Protein of SEQ ID NO:245 (internal designation 105-026-1-0-A5-CS)

The protein of SEQ ID NO:245 encoded by the cDNA of SEQ ID NO:4 displays a leucine zipper pattern situated near its COOH terminal part (position 371 to 392).

It is believed that the protein of SEQ ID NO: 245 is able to dimerize either with itself (homo-dimerisation) or with an heterologous protein (hetero-dimerisation) of interest, through the mediation of its leucine zipper domain. Preferred polypeptides of the invention are polypeptides comprising fragments of SEQ ID NO: 245 from position 371 to 392, and fragments having any of the biological activities described herein.

Protein of SEQ ID NO: 257 (internal designation 106-043-4-0-H3-CS)

The 265-amino-acid-long protein of SEQ ID: 257 encoded by the cDNA of SEQ ID NO: 16 exhibits homology to the Homo sapiens hypothetical protein (Genbank accession number AJ278482). These two proteins are probably the result of an alternative splicing.

5 The protein of SEQ ID NO: 257 displays a leucine zipper pattern situated from position 155 to 176. Thus, it is believed that the protein of SEQ ID NO: 257 is able to dimerize either with itself (homo-dimerisation) or with an heterologous protein (hetero-dimerisation) of interest, through the mediation of its leucine zipper domain. Preferred polypeptides of the invention are polypeptides comprising leucine zipper domains fragments and fragments having any of the biological activities
10 described herein.

Protein of SEQ ID NO: 314 (internal designation 188-41-1-0-B8-CS.cor)

A growing number of proteins have been shown to undergo post-translational modification by fatty acids that are covalently linked to cysteine residues through a thioester bond. Fatty acid modifications contribute to intracellular protein localization by facilitating membrane binding and
15 also by strengthening protein-protein interactions. Cycles of palmitoylation and depalmitoylation have been described for a number of intracellular proteins, but the relevant enzymes that catalyze these processes have yet to be fully characterized and the full significance of these cycles remains to be elucidated.

Palmitoyl-protein thioesterase-1 (PPT1) is a lysosomal hydrolase that removes long-chain
20 fatty acyl groups from modified cysteine residues in proteins. Mutations in PPT1 have been found to cause the infantile form of neuronal ceroid lipofuscinosis (INCL).

Soyombo and Hofmann (J. Biol. Chem. 272: 27456-27463 [1997]) identified cDNAs encoding PPT2. The deduced PPT2 protein contains 302 amino acids, including a 27-amino acid leader peptide, a sequence motif characteristic of many thioesterases and lipases, and 5 potential N-
25 linked glycosylation sites. PPT2 shares 18% amino acid identity with PPT1. Soyombo and Hofmann tentatively localized the human PPT2 gene to 6p21.3. Northern blot analysis detected a predominant 2.0-kb PPT2 transcript in the human tissues examined, with the highest expression in skeletal muscle; variable amounts of 2.8- and 7.0-kb transcripts were also observed.

Cell fractionation studies indicate that PPT2 is present in the lysosomal fraction.
30 Immunoblot analysis of recombinant PPT2 expressed in mammalian cells showed 6 PPT2 proteins ranging in size from 31 to 42 kDa. Treatment that removes asparagine-linked oligosaccharides resulted in a single major protein of 31 kDa and a minor protein of 33 kDa.

Recombinant PPT2, like PPT1, possesses thioesterase activity and localizes to the lysosome. Since PPT2 could not substitute for PPT1 in correcting the metabolic defect in INCL
35 cells and was unable to remove palmitate groups from palmitoylated proteins, it appears that PPT2 possesses a different substrate specificity than PPT1. Another study, however, was able to show,

after expression of the recombinant protein in a baculovirus system and using cell lysate as substrate, that the protein had S-thioesterase activity with a preference for acyl groups palmitic and myristic acid.

The subject invention provides the protein/polypeptide of SEQ ID NO:314, encoded by the cDNA of SEQ ID NO:73. The invention also provides biologically active fragments of SEQ ID NO:314. In one embodiment, the polypeptides of SEQ ID NO:314 are interchanged with the corresponding polypeptide encoded by the human cDNA of clone 188-41-1-0-B8-CS. "Biologically active fragments" are defined as those peptide or polypeptide fragments having at least one of the biological functions of the full length protein (e.g., removal of long-chain fatty acyl groups from modified cysteine residues in proteins). Compositions of the protein/polypeptide of SEQ ID NO:314, or biologically active fragments thereof, are also provided by the subject invention. These compositions may be made according to methods well known in the art.

The invention also provides variants of the protein of SEQ ID NO:314. These variants have at least about 80%, more preferably at least about 90%, and most preferably at least about 95% amino acid sequence identity to the amino acid sequence encoded by SEQ ID NO:73. Variants according to the subject invention also have at least one functional or structural characteristic of the protein of SEQ ID NO:314. The invention also provides biologically active fragments of the variant proteins. Compositions of variants, or biologically active fragments thereof, are also provided by the subject invention. These compositions may be made according to methods well known in the art. Unless otherwise indicated, the methods disclosed herein can be practiced utilizing the protein encoded by SEQ ID NO:73, biologically active fragments of SEQ ID NO:314, variants of SEQ ID NO:314, and biologically active fragments of the variants.

Because of the redundancy of the genetic code, a variety of different DNA sequences can encode the amino acid sequence of SEQ ID NO:314. In a preferred embodiment, SEQ ID NO:314 is encoded by clone 188-41-1-0-B8-CS or the cDNA of SEQ ID NO:73. It is well within the skill of a person trained in the art to create these alternative DNA sequences which encode proteins having the same, or essentially the same, amino acid sequence. These variant DNA sequences are, thus, within the scope of the subject invention. As used herein, reference to "essentially the same" sequence refers to sequences that have amino acid substitutions, deletions, additions, or insertions that do not materially affect biological activity. Fragments retaining one or more characteristic biological activity of the protein encoded by clone 188-41-1-0-B8-CS are also included in this definition.

In one aspect of the subject invention, SEQ ID NO:314, and variants thereof, can be used to generate polyclonal or monoclonal antibodies. Both biologically active and immunogenic fragments of SEQ ID NO:314, or variant proteins, can be used to produce antibodies. Polyclonal and/or monoclonal antibodies can be made according to methods well known to the skilled artisan.

Antibodies produced in accordance with the subject invention can be used in a variety of detection assays known to those skilled in the art.

SEQ ID NO:314 can be used as a marker for identification of lysosome dysfunction in individuals. In this aspect of the subject invention, antibodies specific for SEQ ID NO:314, or
5 fragments thereof, are used in routine immunoassays to screen for the presence or absence of SEQ ID NO:314, or fragments thereof, in samples containing lysosomal contents. The presence or absence of the protein of SEQ ID NO:314 can be used to provide an indication of lysosomal function and is, thus, useful for diagnostic/prognostic identification of lysosomal dysfunction.

The subject invention also provides materials and methods for the screening of individual
10 samples for the presence or absence of nucleic acids encoding the protein of SEQ ID NO:314, or variants thereof. In one embodiment, nucleic acids are provided for hybridization assays, known to those skilled in the art, of mRNA or cDNA. The hybridization assays are performed upon nucleic acid samples obtained, or derived from, an individual with suspected lysosomal dysfunction. The hybridization assays screen for the presence or absence of nucleic acids encoding SEQ ID NO:314,
15 or variants thereof. The presence or absence of such nucleic acids can be used as a predictive/prognostic indicator of disease state or lysosome function.

Nucleic acids of the invention can also be used in gene replacement or gene therapy protocols. This aspect of the subject invention nucleic acids encoding SEQ ID NO:314, or biologically active fragments thereof, can be introduced into cells and implanted into an individual
20 with lysosomal disorders. In one embodiment, genetically engineered macrophage can be used for the treatment regimen (see, for example, Eto and Ohashi [2000] J. Inherit. Metabol. Dis. 23:293-298). Alternatively, autologous cells may be obtained from an individual, transformed with nucleic acid *ex vivo*, expanded *ex vivo*, and reintroduced into the individual. Such methods are well known to the skilled artisan.

25 *Protein of SEQ ID NO:280 (internal designation 160-75-4-0-A9-CS):*

The protein of SEQ ID NO:280, encoded by the cDNA of SEQ ID NO:39 and expressed in the fetal brain, is a chromosome 12 paralog of C7orf2, a human protein described as a transmembrane receptor located on chromosome 7 (Heus, H. C., A. Hing, et al. (1999) Genomics 57(3): 342-51). In addition, this protein is an ortholog of the murine gene LMBR1L, found to be
30 involved in polydactily in mice (Clark, R. M., P. C. Marker, et al. (2000) Genomics 67(1): 19-27). A high level of homology was also found with a gene identified in Fugu rubripes (AF056116), as well as with C. Elegans R05D3.2 (Gellner, K. and S. Brenner (1999) Genome Res 9(3): 251-8).

The 362-amino-acid-long protein of SEQ ID NO:280, encoded by the cDNA of SEQ ID NO:39 is a splice variant of Z64989, located on chromosome 12. The chromosome 12 gene has 6
35 known variants described in entries AK001356 and AK001651 in genbank and entries A26354, A26375, X27360 and Z64989 in geneseqn. The closest sequence is Z64989, either at the nucleotide

or the protein level. Z64989 is split into 17 exons, of which the protein of the invention contains the last 14. The transcription start site of the cDNA of SEQ ID NO:39 lies within the third intron of Z64989, and the protein of the invention starts at position 128 of Z64989. In addition, 2 potential leucine zippers are present in the protein of the invention (positions 136-157 and 272-293).

5 Preaxial polydactyly is a congenital hand malformation that includes duplicated thumbs, various forms of triphalangeal thumbs, and duplications of the index finger. Clark et al. (supra) demonstrated the correspondence between the spatial and temporal changes in *Lmbr1* expression and the embryonic onset of polydactyly mutant phenotype, suggesting that a downregulation of *Lmbr1* results in polydactyly. It is likely that the *Lmbr1* gene is involved in the patterning of limbs
10 during mammalian development, for example by receiving and transducing a locally secreted ligand in the developing limb.

 It is believed that the protein of SEQ ID NO:280 is a paralog of human C7orf2, and is thus a membrane bound protein implicated in the patterning of the mammalian body plan during early development. For example, the protein of the invention may be involved in organizing limb
15 development, as well as in the development of the fetal brain. As such, the activity of the present protein likely influences various cellular processes, including gene expression, cellular growth and proliferation, as well as cellular differentiation. In addition, leucine zippers within the present protein render the protein capable of undergoing specific protein-protein interactions with other leucine-zipper containing proteins, including with itself (i.e. homodimerization). Preferred
20 polypeptides of the invention are fragments of SEQ ID NO:280 having any of the biological activities described herein.

 In one embodiment of the present invention, the present protein can be used to identify cells of the fetal brain. For example, the protein of the invention or part thereof may be used to synthesize specific antibodies using any technique known to those skilled in the art. Such tissue-
25 specific antibodies may then be used to identify tissues of unknown origin, such as in forensic samples, differentiated tumor tissue that has metastasized to foreign bodily sites, etc., or to differentiate different tissue types in a tissue cross-section using immunochemistry. In addition, labeled reagents that can specifically bind to the protein of the invention can be used to visualize cell membranes and the components of the secretory pathway in cells, e.g. the ER and Golgi.

30 In another embodiment of the present invention, the present protein can be used to diagnose developmental abnormalities, or the potential for such abnormalities, e.g. in a fetus or in adults to determine (i.e. to determine if they are a carrier of a mutant copy of the gene). Individuals found to carry one or two mutant copies of the present gene would be candidates for, e.g. gene therapy or other strategies to correct or compensate for the gene deficiency, or for strategies to
35 ensure that their children would not be carriers of the mutated gene. The characterization of mutations in genes encoding the present protein would also be of great value in understanding the

nature of polydactyly and other developmental disorders, thereby facilitating the development of other strategies for treating and preventing these disorders.

In another embodiment, the present protein is used to modulate gene expression, cell growth and proliferation, and/or cell differentiation in cells in vitro or in vivo. For example, any of these behaviors can be increased or inhibited in cells grown in vitro, e.g. for protein production or for ex vivo therapeutic strategies. In addition, any disease associated with an increase or decrease in any of these cellular behaviors in vivo can be treated or prevented by enhancing or inhibiting the expression or activity of the protein of the invention in cells in vivo.

Proteins of SEQ ID NOs: 309 and 304 (internal designations 188-11-1-0-B3-CS and 187-34-0-0-112-CS)

The proteins of SEQ ID NOs: 309 and 304 are encoded by the cDNAs of SEQ ID NOs: 68 and 63. Accordingly, it will be appreciated that all characteristics and uses of the polypeptides of SEQ ID NOs: 309 and 304 described throughout the present application also pertain to the polypeptides encoded by human cDNA of clones 188-11-1-0-B3-CS and 187-34-0-0-112-CS. In addition, it will be appreciated that all characteristics and uses of the nucleic acids of SEQ ID NOs: 68 and 63 described throughout the present application also pertain to the nucleic acids of the human cDNAs of clones 188-11-1-0-B3-CS and 187-34-0-0-112-CS.

The protein of SEQ ID NO: 309 (encoded by the clone having internal designation number 188-11-1-0-B3-CS) and the polymorphic variant thereof of SEQ ID NO: 304 (encoded by the clone having internal identification number 187-34-0-0-112-CS and which differs from the polypeptide encoded by the clone having internal designation number 188-11-1-0-B3CS at a single amino acid), are highly homologous to the first 279 amino acids of the LGI1 (Leucine-rich gene – Glioma Inactivated) protein. Clones 188-11-1-0-B3-CS and 187-34-0-0-112-CS appear to be splicing and polymorphic variants of LGI1. The LGI1 protein is 557 amino acid in length. (See Somerville et al., (2000) Mammalian Genome 11, 622-627 ; Chernova, et al. (1998) Oncogene 17, 2873-2881, the disclosures of which are incorporated herein by reference in their entireties). Clone 188-11-1-0-B3-CS align with the first 279 amino acids of LGI1, followed by the addition of 12 amino acids (VLREIHRFTNMS) to the C-terminal end which do not appear to be homologous to LGI1. Like LGI1, clone 188-11-1-0-B3-CS and the polymorphic variant 187-34-0-0-112-CS contain the LRR domain and are highly expressed in brain tissue.

LGI1 belongs to a large family of leucine-rich repeat (LRR) proteins. It is believed that the LRR domains act as a region of protein-protein interaction. This has been substantiated as the family of known LRR proteins has grown. Leucine-rich repeats have been identified as essential components in glycoprotein hormone receptors, proteoglycans and the Trk proteins by expression of mutants and artificial chimaeras in tissue culture and by biochemical analysis of the properties of these constructs. Many transmembrane LRR proteins are known or suspected to encode truncated

forms (N and L⁶, and slit for example) with functional significance. The proteoglycan Decorin, a secreted protein, binds TGF- β , a growth factor which stimulates decorin expression. Since decorin inhibits growth of cultured cells, it may form part of a negative feedback loop to regulate cell growth. This is similar to the proposed function of the LGI1 receptor protein.

5 Analysis of brain gliomas has revealed that LGI1 expression is either abolished or greatly reduced in high-grade tumors compared with more benign ones, indicating a role as a tumor suppressor gene (Cowell et al. 2000; Cowell et al. 1998, the disclosure of which is incorporated herein by reference in its entirety). Most glioblastoma multiforme (GBM) brain tumors contain only one genomic copy of LGI1, and this one is almost invariably not expressed. How the gene is
10 inactivated is not clear, although one possibility is that chromosome or gene rearrangement, which occur in 20-25% of tumors, cause inactivation as a result of a positional effect. Recently it was determined that the LGI1 gene is located on 10q24, and is disrupted by translocation in the T98G GBM cell line and is also rearranged in over 26% of primary brain tumors. Alternatively, LGI1 may be part of a highly regulated pathway where inactivation of other key members or high specific
15 transcription factors results in either inactivation of all genes in the pathway or a failure to initiate transcription.

 Since functional inactivation of LGI1 occurs during the transition of low-grade to high-grade brain tumors, knockout or transgenic mice in which the expression of the protein of SEQ ID NO:309 or 304 has been reduced, eliminated or altered may be used as disease model. In particular,
20 mice that overexpress LGI1 may be used as a tumorigenesis model.

 Mice are particularly useful as models for assessing the consequences of altering the level or activity of the proteins of SEQ ID NO:309 or 304 or to identify agents useful in treating tumorigenesis, since human and mouse LGI1 are highly conserved, showing 91% identity at the nucleotide level and 97% similarity at the amino acid level, with most of the amino acid
25 substitutions being conservative. The mouse *lgi1* gene is 4.2 kb in length, while the human LGI1 is 2.2 kb in length. This difference in size between the human and mouse gene is a result of the inclusion of a 2 kb sequence in the 3' untranslated region in the mouse gene. Whether the additional sequence affects gene expression is not clear. Further analysis of the genomic sequence reveals that the number of exon/intron boundaries is also similar in humans and mice. The high
30 degree of LGI1 conservation between mice and humans implies that this gene has experienced a strong selection pressure. It is intriguing to speculate that any major deviations in the primary protein sequence may result in a loss of function of this gene product. Total or partial loss of the LGI1 gene function could, therefore, be lethal, which in turn implies that LGI1 plays an important role in normal brain development as well as in tumor formation.

35 SEQ ID NOs:309 and 304 also have high homology with Slit, a secreted *Drosophila* protein which plays a role in the development of axon pathway development in the central nervous system. The Slit protein is necessary for the normal development of the midline on the CNS, particularly the

midline glial cells, and for the concomitant formation of the commissural axon pathway. The process is dependent on the level of Slit protein expression. It appears that the Slit protein is excreted by the midline glial cells, where it is synthesized and is eventually associated with the surface axons that traverse them. Contact of cells with supernatant expressing the product of this gene increases the permeability of THP-1 monocyte cells to calcium. Thus, it is likely that Slit is involved in a signal transduction pathway that is initiated when Slit protein binds a receptor on the surface of the monocyte cell.

In view of the above, it is believed that the proteins of SEQ ID NOs:309 and 304 are involved in a signal transduction pathway mediated through a receptor that modulates the differentiation and/or proliferation of cells.

Northern blot analysis detects LGI1 transcripts only in brain, neural tissue, and skeletal muscle but not in heart, kidney, lung, placenta, liver, or pancreas. Northern blot analysis of RNA derived from several different regions of human brain revealed a widespread expression of LGI1 although with different intensities. The highest abundance was found in cerebral cortex, hippocampus, and putamen. The lowest expression was detected in corpus callosum. The levels of expression were intermediate in the other brain regions. Accordingly, the proteins of SEQ ID NOs:309 or 304 or fragments thereof, as well as polynucleotides encoding the proteins of SEQ ID NOs:309 or 304, may be used to determine whether a tissue sample is derived from brain (and in particular cerebral cortex, hippocampus, or putamen), neural tissue, and skeletal tissue or to distinguish whether a tissue sample is derived from brain or another tissue, such as heart, kidney, lung, placenta, liver, or pancreas.

Accordingly, the present invention includes the use of the protein of SEQ ID NOs: 309 or 304, fragments comprising at least 5, 8, 10, 12, 15, 20, 25, 30, 35, 40, 50, 60, 75, 100, 150, or 200 consecutive amino acids thereof, or fragments having a desired biological activity to treat or ameliorate a condition, such as those listed above, in an individual. In such embodiments, the protein of SEQ ID NO:309 or 304, or a fragment thereof, is administered to an individual in whom it is desired to increase or decrease any of the activities of the protein of SEQ ID NO:309 or 304, including tumor suppression, modulation of neural development or involvement in brain tumors, glioblastoma multiforme, brain injuries, neurodegenerative disease states and behavioral disorders such as Alzheimers Disease, Parkinsons Disease, epilepsy, multiple sclerosis, Huntingtons Disease, schizophrenia, obsessive compulsive disorders, and in the processes of nerve regeneration in spinal cord injury, stroke, facial nerve damage, diabetes caused nerve damage, and retinal regeneration.

The protein of SEQ ID NO:309 or 304 or a fragment thereof may be administered directly to the individual or, alternatively, a nucleic acid encoding the protein of SEQ NO:309 or 304 or a fragment thereof may be administered to the individual. Alternatively, an agent which increases the activity of the protein of SEQ ID NO:309 or 304 may be administered to the individual. Such agents may be identified by contacting the protein of SEQ NO:309 or 304 or a cell or preparation

containing the protein of SEQ ID NO:309 or 304 with a test agent and assaying whether the test agent increases the activity of the protein. For example, the test agent may be a chemical compound or a polypeptide or peptide.

Alternatively, the activity of the protein of SEQ ID NO:309 or 304 may be decreased by administering an agent which interferes with such activity to an individual. Agents which interfere with the activity of the protein of SEQ ID NO:309 or 304 may be identified by contacting the protein or a cell or preparation containing the with a test agent and assaying whether the test agent decreases the activity of the protein. For example, the agent may be a chemical compound, a polypeptide or peptide, an antibody, or a nucleic acid such as an antisense nucleic acid or a triple helix-forming nucleic acid.

In one embodiment, the invention relates to methods and compositions using the protein of the invention or part thereof as a marker protein to selectively identify tissues, preferably brain, or to distinguish between two or more possible sources of a tissue sample on the basis of the level of the protein of SEQ ID NO:309 or 304 in the sample. For example, the protein of SEQ ID NO:309 or 304 or fragments thereof may be used to generate antibodies using any techniques known to those skilled in the art, including those described therein. Such tissue-specific antibodies may then be used to identify tissues of unknown origin, for example, forensic samples, differentiated tumor tissue that has metastasized to foreign bodily sites, or to differentiate different tissue types in a tissue cross-section using immunochemistry. In such methods a tissue sample is contacted with the antibody, which may be detectably labeled, under conditions which facilitate antibody binding. The level of antibody binding to the test sample is measured and compared to the level of binding to control cells from brain or tissues other than brain to determine whether the test sample is from brain. Alternatively, the level of the protein of SEQ ID NO:309 or 304 in a test sample may be measured by determining the level of RNA encoding the protein of SEQ ID NO:309 or 304 in the test sample. RNA levels may be measured using nucleic acid arrays or using techniques such as in situ hybridization, Northern blots, dot blots or other techniques familiar to those skilled in the art. If desired, an amplification reaction, such as a PCR reaction, may be performed on the nucleic acid sample prior to analysis. The level of RNA in the test sample is compared to RNA levels in control cells from brain or tissues other than brain to determine whether the test sample is from brain. For a number of disorders listed above, particularly of the nervous system, expression of the genes encoding the polypeptide of SEQ ID NO:309 or 304 at significant higher or lower levels may be routinely detected in certain tissues or cell types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, synovial fluid, and spinal fluid) or another tissue of cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

In another embodiment, antibodies to the protein of SEQ ID NO:309 or 304 or part thereof may be used for detection, enrichment, or purification of cells expressing the protein of SEQ ID

NO:309 or 304, including using methods known to those skilled in the art. For example, an antibody against the protein of SEQ ID NO:309 or 304 or a fragment thereof may be fixed to a solid support, such as a chromatography matrix. A preparation containing cells expressing the protein of SEQ ID NO:309 or 304 is placed in contact with the antibody under conditions which facilitate
5 binding to the antibody. The support is washed and then the cells are released from the support by contacting the support with agents which cause the cells to dissociate from the antibody.

In another embodiment of the present invention, the protein of SEQ ID NO:309 or 304 or a fragment thereof may be used to diagnose disorders associated with altered expression of the protein of SEQ ID NO:309 or 304. In some embodiments, the protein of SEQ ID NO:309 or 304 or
10 fragments thereof may be used to diagnose cancer. In such techniques, the level of the protein of SEQ ID NO:309 or 304 in an ill individual is measured using techniques such as those described herein and compared to the level in normal individuals. For example, a decreased level of the protein of SEQ ID NO:309 or 304 relative to normal individuals suggests that the ill individual may suffer from cancer or be predisposed to getting cancer in the future.

Another embodiment of the present invention is a polypeptide comprising a structural or functional domain of the protein of SEQ ID NO:309 or 304. Such structural or functional domains of the protein of SEQ ID NO:309 or 304 include a leucine rich repeat C-terminal domain located between amino acid positions 173 and 222, a leucine rich repeat located between amino acid
15 positions 92 and 115, a leucine rich repeat located between amino acid positions 116 and 139, a leucine rich repeat located between amino acid positions 140 and 163, a leucine rich repeat located between amino acid positions 164 and 185, a membrane spanning segment located between amino acid positions 15 and 35, and a signal peptide comprising the sequence FLCLLSALLLTEG/KK.
20

Accordingly, the protein of SEQ ID NO:309 or 304 or fragments thereof, or polynucleotides encoding these proteins or fragments, may be used in *in vitro* diagnostic assays for malignant brain
25 tumors, such as glioblastoma multiforme. These proteins or nucleic acids may also be used in the attenuation / prevention and/or treatment of brain tumors and/or brain injuries, of neurodegenerative disease states and behavioral disorders such as Alzheimers Disease, Parkinsons Disease, epilepsy, multiple sclerosis, Huntingtons Disease, schizophrenia, obsessive compulsive disorders, and in the processes of nerve regeneration in spinal cord injury, stroke, facial nerve damage, diabetes caused
30 nerve damage, and retinal regeneration.

In addition, the protein, as well as, antibodies directed against the protein, and relevant small molecules may be used as tumor markers and /or immunotherapy targets for the above disease states. For example, antibodies directed against amino acids VLREIHRFTNMS of both clones may aid in the differential detection of the secreted and receptor forms of this protein, since the proteins
35 of SEQ ID NOs:309 and 304 have homology to the secreted forms of LGI1. In addition, the proteins of SEQ ID NOs:309 and 304 or fragments thereof may be used to identify binding partners as described herein.

DNA-binding proteins

The invention relates to compositions and methods using proteins of the invention containing a DNA-binding domain, herein referred to as DBP, such as the ones described in this section and those containing a DNA binding domain domain as shown on Table VI, or parts thereof, preferably
5 fragments comprising a DNA binding domain, or derivative thereof.

Transcriptional regulation is primarily achieved by the sequence-specific binding of proteins to DNA and RNA. Of the known protein motifs involved in the sequence specific recognition of DNA, the zinc finger protein is unique in its modular nature. Zinc finger domains are found in numerous zinc binding proteins which are involved in protein-nucleic acid interactions. They are independently folded
10 zinc-containing mini-domains which are used in a modular repeating fashion to achieve sequence-specific recognition of DNA (Klug 1993 Gene 135, 83-92). Such zinc binding proteins are commonly involved in the regulation of gene expression, and usually serve as transcription factors (see US patents 5,866,325; 6,013,453 and 5,861,495).

To date, zinc finger proteins have been identified which contain between 2 and 37 modules.
15 More than two hundred proteins, many of them transcription factors, have been shown to possess zinc fingers domains. Zinc fingers connect transcription factors to their target genes mainly by binding to specific sequences of DNA. Zinc finger modules are found in a wide variety of transcription regulatory proteins in eukaryotic organisms. A zinc finger domain is generally composed of 25 to 30 amino acid residues which form one or more tetrahedral ion binding sites. The binding sites contain four ligands
20 consisting of the sidechains of cysteine, histidine and occasionally aspartate or glutamate. The binding of zinc allows the relatively short stretches of polypeptide to fold into defined structural units which are well-suited to participate in macromolecular interactions (Berg, J. M. et al. (1996) Science 271:1081-1085). The zinc finger domain was first recognized in the transcription factor TFIIIA from *Xenopus* oocytes (Miller, et al., EMBO, 4:1609-1614, 1985; Brown, et al., FEBS Lett., 186:271-274, (1985)).

25 Zinc binding domains which contain a C_3HC_4 sequence motif are known as RING domains (Lovering, R. et al. (1993) Proc. Natl. Acad. Sci. USA 90:2112-2116). The RING domain consists of eight metal binding residues, and the sequences that bind the two metal ions overlap (Barlow, P. N. et al. (1994) J. Mol. Biol. 237:201-211). Functions of RING finger proteins are mediated through DNA binding and include the regulation of gene expression, DNA recombination, and DNA repair
30 (see Borden and Freemont, Curr Opin Struct Biol 6:395-401 (1996) and US patent 5,861,495).

Both the RING finger and the LIM domain mediate protein-protein interactions and are involved in transcriptional control, either by directly affecting transcription or recruiting co-activators or co-repressors. LIM domains also contribute to various signalling pathways. They may interact with protein kinases and anchor gene products to large protein complexes or to cellular compartments.

35 PHD fingers are C_4HC_3 zinc fingers spanning approximately 50-80 residues and distinct from RING fingers or LIM domains. They are thought to be mostly DNA or RNA binding domain but may also be involved in protein-protein interactions (for a review see Aasland et al, Trends Biochem Sci

20:56-59 (1995)). The PHD finger domain, belonging to zinc finger domain family, is found in many regulatory proteins which are frequently associated with chromatin-mediated transcriptional regulation.

The nucleic acid binding activity of DBP or part thereof may be assayed using any of the
5 assays known to those skilled in the art including those described in US patent 6,013,453.

The invention relates to compositions and methods using DBPs or part thereof, especially fragments comprising a DNA-binding domain, to stimulate gene transcription.

One of the remarkable features of activation domains of transcriptional factors in general is that "fusing" them to heterologous protein domains seldom affects their ability to activate transcription
10 when recruited to a wide variety of promoters. The high degree of functional independence exhibited by these activation domains makes them valuable tools in various biological assays for analyzing gene expression and protein--protein or protein-RNA or protein-small molecule drug interactions. Several strategies to improve the potency of activation domains and thereby the expression of genes under their control have been reported. These approaches generally involve increasing the number of copies of
15 activation domains fused to the DNA binding domain or generating activators containing synergizing combinations of activation domains.

Therefore, in an additional embodiment, this invention provides compositions and methods containing new transcription factors comprising DBP or part thereof, preferably fragments containing DNA-binding domains. Such transcription factors may be designed to regulate the expression of target
20 genes of interest. Aspects of the invention are applicable to systems involving either covalent or non-covalent linking of the transcription activation domain to a DNA binding domain. In practice, cells can be engineered by the introduction of recombinant nucleic acids encoding the fusion proteins containing at least two mutually heterologous domains, one of them being the DNA-binding domain of the invention, and in some cases additional nucleic acid constructs, to render them capable of ligand-
25 dependent regulation of transcription of a target gene. Administration of the ligand to the cells then regulates (positively, or in some cases, negatively) target gene transcription (all laboratory methods related to this embodiment are completely described in US patents 6.015.709, which disclosure is hereby incorporated by reference in its entirety). Illustrative (non-limiting) example of heterologous domains which can be included along with a DNA-binding domain in various fusion proteins of this
30 invention include another transcription regulatory domains (i.e., transcription activation domains such as a p65, VP16 or AP domain; transcription potentiating or synergizing domains; or transcription repression domains such as an ssn-6/TUP-1 domain or Kruppel family suppressor domain); a DNA binding domain such as a GAL4, lex A or a composite DNA binding domain such as a composite zinc finger domain or a ZFHD1 domain; or a ligand-binding domain comprising or derived from (a) an
35 immunophilin, cyclophilin or FRB domain; (b) an antibiotic binding domain such as tetR; or (c) a hormone receptor such as a progesterone receptor or ecdysone receptor. A wide variety of ligand binding domains may be used in this invention, although ligand binding domains which bind to a cell

permeant ligand are preferred. It is also preferred that the ligand have a molecular weight under about 5 kD, more preferably below 2.5 kD and optimally below about 1500 D. Non-proteinaceous ligands are also preferred. Examples of ligand binding domain/ligand pairs that may be used in the practice of this invention include, but are not limited to: FKBP:FK1012, FKBP:synthetic divalent FKBP ligands (see 5 WO 96/0609 and WO 97/31898), FRB:rapamycin/FKBP (see e.g., WO 96/41865 and Rivera et al, "A humanized system for pharmacologic control of gene expression", *Nature Medicine* 2(9):1028-1032 (1997)), cyclophilin:cyclosporin (see e.g. WO 94/18317), DHFR:methotrexate (see e.g. Licitra et al, 1996, *Proc. Natl. Acad. Sci. U.S.A.* 93:12817-12821), TetR:tetracycline or doxycycline or other analogs or mimics thereof (Gossen and Bujard, 1992, *Proc. Natl. Acad. Sci. U.S.A.* 89:5547; Gossen et al, 10 al, 1995, *Science* 268:1766-1769; Kistner et al, 1996, *Proc. Natl. Acad. Sci. U.S.A.* 93:10933-10938), a progesterone receptor:RU486 (Wang et al, 1994, *Proc. Natl. Acad. Sci. U.S.A.* 91:8180-8184), ecdysone receptor:ecdysone or muristerone A or other analogs or mimics thereof (No et al, 1996, *Proc. Natl. Acad. Sci. U.S.A.* 93:3346-3351) and DNA gyrase:coumermycin (see e.g. Farrar et al, 1996, *Nature* 383:178-181). In many applications it is preferable to use a DNA binding domain which 15 is heterologous to the cells to be engineered. In the case of composite DNA binding domains, component peptide portions which are endogenous to the cells or organism to be engineered are generally preferred.

In another aspect of this embodiment, polynucleotides encoding DNA-binding domains as well as any other functional fragments of DBP may be introduced into polynucleotides encoding fusion 20 proteins for a variety of regulated gene expression systems, including both allostery-based systems such as those regulated by tetracycline, RU486 or ecdysone, or analogs or mimics thereof, and dimerization-based systems such as those regulated by divalent compounds like FK1012, FKCsA, rapamycin, AP1510 or coumermycin, or analogs or mimics thereof, all as described below (See also, Clackson, Controlling mammalian gene expression with small molecules, *Current Opinion in Chem. Biol.* 1:210-25 218 (1997)). The fusion proteins may comprise any combination of relevant components, including bundling domains, DNA binding domains, transcription activation (or repression) domains and ligand binding domains. Other heterologous domains may also be included.

Another embodiment of this invention relates to expression systems, preferably vectors and vector-containing cells, using DBP or part thereof, especially the DNA-binding domain. In this regard, 30 recombinant nucleic acids are provided which encode fusion proteins containing the transcription activation domain of the invention and at least one additional domain that is heterologous thereto, where the peptide sequence of said activation domain is itself eventually modified relative to the naturally occurring sequence from which it was derived to increase or decrease its potency as a transcriptional activator relative to the counterpart comprising the native peptide sequence. Each of the 35 recombinant nucleic acids of this invention may further comprise an expression control sequence operably linked to the coding sequence and may be provided within a DNA vector, e.g., for use in transducing prokaryotic or eukaryotic cells. Some of the recombinant nucleic acids of a given

composition as described above, including any optional recombinant nucleic acids, may be present within a single vector or may be apportioned between two or more vectors. The recombinant nucleic acids may be provided as inserts within one or more recombinant viruses which may be used, for example, to transduce cells in vitro or cells present within an organism, including a human or non-human mammalian subject. It should be appreciated that non-viral approaches (naked DNA, liposomes or other lipid compositions, etc.) may be used to deliver recombinant nucleic acids of this invention to cells in a recipient organism. The resultant engineered cells and their progeny containing one or more of these recombinant nucleic acids or nucleic acid compositions of this invention may be used in a variety of important applications, including human gene therapy, analogous veterinary applications, the creation of cellular or animal models (including transgenic applications) and assay applications. Such cells are useful, for example, in methods involving the addition of a ligand, preferably a cell permeant ligand, to the cells (or administration of the ligand to an organism containing the cells) to regulate expression of a target gene.

The invention also relates to methods and compositions using DBP or part thereof to bind to nucleic acids, preferably DNA, alone or in combination with other substances. For example, DBP or part thereof is added to a sample containing nucleic acid in conditions allowing binding, and allowed to bind to nucleic acids. In a preferred embodiment, DBP or part thereof may be used to purify nucleic acids such as restriction fragments. In another preferred embodiment, DBP or part thereof may be used to visualize nucleic acids when the polypeptide is linked to an appropriate fusion partner, or is detected by probing with an antibody. Alternatively, DBP or part thereof may be bound to a chromatographic support, either alone or in combination with other DNA binding proteins, using techniques well known in the art, to form an affinity chromatography column. A sample containing nucleic acids to purify is run through the column. Immobilizing DBP or part thereof on a support advantageous is particularly for those embodiments in which the method is to be practiced on a commercial scale. This immobilization facilitates the removal of the protein from the batch of product and subsequent reuse of the protein. Immobilization of DBP or part thereof can be accomplished, for example, by inserting a cellulose-binding domain in the protein. One of skill in the art will understand that other methods of immobilization could also be used and are described in the available literature.

In another embodiment, the present invention relates to compositions and methods using DBP or part thereof, especially the DNA-binding domain, to alter the expression of genes of interest in a target cells. Such genes of interest may be disease related genes, such as oncogenes or exogenous genes from pathogens, such as bacteria or viruses using any techniques known to those skilled in the art including those described in US patents 5,861,495; 5,866,325 and 6,013,453.

In still another embodiment, DBP or part thereof may be used to diagnose, treat and/or prevent disorders linked to dysregulation of gene transcription such as cancer and other disorders relating to abnormal cellular differentiation, proliferation, or degeneration, including

hyperaldosteronism, hypocortisolism (Addison's disease), hyperthyroidism (Grave's disease), hypothyroidism, colorectal polyps, gastritis, gastric and duodenal ulcers, ulcerative colitis, and Crohn's disease.

Protein of SEQ ID NO: 388 (internal designation 109-002-4-0-C6-CS)

5 The protein of SEQ ID NO: 388 encoded by cDNA of SEQ ID NO: 147 is a 375 amino-acids long protein containing a zinc finger domain, namely a PHD-finger domain from positions 329 to 339.

 The PHD finger was originally identified by comparison of the maize homeodomain (HD) protein ZMHOX1a (Bellmann R. and Werr W. EMBO J. 11: 3367-3374 (1992)) to its Arabidopsis relative HAT3.1 and named plant homeodomain (PHD) finger due to its association with the DNA-
10 binding HD in both genes. This motif often occurs in various regulatory genes, such as members of the trithorax (TRX-G) or polycomb (PC-G) groups (Aasland R. et al. Trends Biochem.Sci. 20: 56-59 (1995)) and leukaemia-associated proteins (LAP finger) (Saha V. et al. Proc.Natl.Acad.Sci. USA 92: 9737-9741 (1995)). The established function of TRX-G and PC-G genes in chromatin modulation in *Drosophila* led to the suggestion that the PHD finger is involved in chromatin-mediated transcriptional
15 control. Recent data provide evidence that PHD finger proteins are associated with chromatin remodelling complexes (Bochar D.A. et al. Proc.Natl.Acad.Sci. USA 97: 1038-1043 (2000)) or contribute to histone acetylation (Loewith R. et al. Mol.Cell.Biol. 20: 3807-3816 (2000)). Based on the position of the unique His residue, the cysteine scaffold of the PHD finger (Cys4-His-Cys3) is clearly distinct from RING fingers (Cys3-His-Cys4) and LIM domains (Cys2-His-Cys5) and from DRIL
20 domains, where two RING finger motifs are closely linked. In contrast to the accumulating knowledge about LIM domains, functional data concerning the PHD finger remain rare (see rev. Halbach T. et al. Nucleic Acids Research 28: 3542-3550 (2000)).

 GYMNOS, a recently described member of the SWI2/SNF2 protein family in plants (22), also contains a PHD finger and takes part in the control of development. The second PHD finger motif of
25 *Drosophila* dMI-2 protein (a reference for animal counterparts) shares high sequence conservation to known plant PHD fingers. Due to the similarity to the *Drosophila* MI-2 gene, GYMNOS has been implicated in chromatin modulation. While the PHD finger is an isolated motif in GYMNOS, the characteristic Cys4-His-Cys3 scaffold in PHDf-HD plant genes is embedded in a large region. This region shares 60% identical residues between seven genes of different plant species and is more highly
30 conserved than the HD (40%). This conservation suggests that the PHD finger is part of a larger functional unit. When combined with a leucine zipper in the surrounding conserved 180 amino acid region in the PHDf-HD proteins, PHD finger activity is masked and silenced. The leucine zipper upstream of the PHD finger mediates interactions with helix 4 of plant 14-3-3 proteins, thus identifying PHDf-HD proteins as potential targets of 14-3-3 signalling pathways. The 14-3-3 family of
35 multifunctional proteins is highly conserved between animals, plants and yeast. Due to the dimeric nature of 14-3-3 proteins and their capacity to form homo- and heterodimers, members of the 14-3-3

protein family function as scaffolds promoting association of protein complexes. 14-3-3 proteins are involved in various signalling pathways that include, for example, Raf, BAD, Bcr/Bcr-Abl, KSR (kinase suppressor of Ras), PKC, PI-3 kinase and cdc25C phosphatase. Others enter the nucleus and are associated with DNA-binding complexes. Recent data even indicate contacts to TBP, TFIIB and the human TBP-associated factor hTAF(II)32 (for rev.see Halsbach T., supra).

Recently PHD finger has been shown to activates transcription in yeast, plant and animal cells. Transcriptional activation in animal cells (in the zebrafish embryo as a test system) tested for different PHD fingers seems to be a general feature of the PHD finger motif in eukaryotic cells.

It remains to be elucidated whether the PHD finger directly interacts with a component of the transcription initiation complex or if its positive effect on transcription is mediated via auxiliary protein interactions. Both assumptions, however, involve PHD finger-mediated protein-protein interactions. Surrounding sequences may interfere sterically with accession of the PHD finger and its exposure could eventually depend on binding of a protein partner.

The PHD finger containing proteins appear to be involved in human diseases. Studies on the AIRE gene from humans (Nagamine K. et al. Nat.Genet. 17: 393-398 (1997), Scott H.S. et al. Mol.Endocrinol. 12: 1112-1119 (1998)) have shed more light on the importance of this motif, since all clinically significant mutations in the AIRE gene coincide with alteration in two PHD fingers, resulting in the rare autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED). The presence of PHD fingers in genes up-regulated in leukaemia, associated with the autoimmune disease APECED or participating in euchromatin to heterochromatin modulation, like the TRX-G or PC-G genes, indicates that this motif may be involved in a variety of important cellular events including developmental disorders, tumors and immune diseases. For example, the role of a chromatin structure remodelling in cancer metastasis and tissue carcinogenesis is well documented (Zhang Y. et al. Cell 16: 279-289 (1998); Klugbauer S. and Rabes H.M. Oncogene 29: 4388-4393 (1999)).

It is believed that the protein of SEQ ID NO: 388 or part thereof is a zinc binding protein, preferably able to bind nucleic acids, more preferably a transcription factor. Preferred polypeptides of the invention are polypeptides comprising the amino acids of SEQ ID NO: 388 from positions 329 to 339. Other preferred polypeptides of the invention are fragments of SEQ ID NO: 388 having any of the biological activity described herein.

In one embodiment of the invention, the protein of the invention, or part thereof, or derivative thereof, may be used to a subject to diagnose developmental disorders and/or cell proliferative disorders linked to dysregulation of gene expression mediated by the PHD-finger domain of the protein of the invention. Such disorders include but are not limited to, renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, epilepsy, gonadal dysgenesis, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spinal bifida, and congenital glaucoma, cataract, sensorineural hearing loss, benign tumors, and cancers such as adenocarcinoma; leukemia; melanoma;

lymphoma; sarcoma; and cancers of the bladder, colon, liver, brain, small intestine, large intestine, breast, ovary, kidney, lung, and prostate. Diagnosis may be performed using nucleic acids or antibodies able to detect the expression of the protein of the invention using any technique known to those skilled in the art including Northern blotting, RT-PCR, immunoblotting methods
5 immunohistochemistry, enzyme-linked immunosorbant assay (ELISA) described herein. Quantities of the protein of the invention expressed in subject samples, control and disease from biopsied tissues or body fluids or cell extracts taken from patients are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment, antagonists or inhibitors of the protein of the invention or part thereof
10 may be administered to patients to treat and/or prevent the above referred disorders. Antagonists or inhibitors of transcriptional activators may indeed be used to suppress transcriptional activation in tumor cells. Such antagonists and/or inhibitors may be antibodies specific for the protein of the invention that can be used directly as an antagonist, or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express the protein of the invention.
15 Neutralizing antibodies, (i.e., those which inhibit protein-protein interactions) are especially preferred for therapeutic use. Other methods to inhibit the expression of the protein of the invention include antisense and triple helix strategies as described herein. Other antagonists or inhibitors of the protein of the invention may be produced using methods which are generally known in the art, including the screening of libraries of pharmaceutical agents to identify those which specifically bind the protein of
20 the invention. The protein of the invention, or part thereof, preferably its functional or immunogenic fragments, or oligopeptides related thereto, can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes, between the protein of the invention, or part thereof, or derivative thereof, and the agent
25 being tested, may be measured. Another technique for drug screening which may be used provides for high throughput screening of compounds having suitable binding affinity to the protein of the invention as described in published PCT application WO84/03564.

Protein of SEQ ID NO: 394 (internal designation 157-17-2-0-C1-CS)

The protein of SEQ ID NO: 394 encoded by the extended cDNA SEQ ID NO: 153 contains
30 a myc-type, helix-loop-helix dimerization domain (Prosite PS00038) from amino acid position 13 to 28 and has no adjacent basic domain. Using the Schiffer-Edmundson helical wheel diagram (Schiffer et al. (1967) Biophys.J. 7:121-135), a hypothetical amphipathic alpha helix is predicted between position 53 and position 68. Three hydrophobic amino acids, Val 55, Phe59 and Ile63, are aligned on the same side of the helix to present a hydrophobic interaction surface and three
35 hydrophilic residues (Tyr53, Gln62 and Ser64) are presented on the other side of helix. There is no Proline residue within the stretch to disrupt the continuity of the alpha helix. Thus, these structural

features in the protein of the invention indicates that this protein could be a novel member of the nonbasic "helix-loop-helix" subfamily (HLH) of transcription regulator.

The helix-loop-helix (HLH) family of transcriptional regulators is involved in the control of different cellular differentiation phenomenon such as neurogenesis, haematopoiesis, myogenesis and angiogenesis. The HLH proteins are found in all eukaryotic organisms ranging from yeast
5 *saccharomyces cerevisiae* to human (Reviewed by Massari ME and Murre C. (2000) *Molecular and Cellular Biology*, 20 (2):429-440). The HLH proteins bind DNA as dimers, and different members of HLH family bind either as homodimers or as heterodimers with other members of the family. The presence in a cell of a large repertoire of distinct complexes that can bind to a particular DNA
10 sequence element suggests that competition for DNA binding may play a regulatory role.

Members of the helix-loop-helix (HLH) family of transcriptional regulation proteins share a common structural element, i.e. a stretch of 40-50 amino acids containing two short amphipathic alpha-helices separated by a linker region (the loop) of varying length (Murre C et al. (1989) *Cell* 56:777-783). This element was initially identified as a region of homology among c-myc, the
15 muscle determination gene MyoD (Davis RL et al. (1987) *Cell* 51:987-1000) and the *Drosophila* achaete-scute complex (AS-C) involved in neural determination (Villares R. and Cabrera CV (1987) *Cell* 50:415-424). The HLH proteins form both homodimers and heterodimers by means of interaction between the hydrophobic residues on the corresponding faces of the two helices to give a parallel four-helix bundle structure (Adrian R et al. (1993) *Nature*, 363:38-45; Ellenberger T et al.
20 (1994) *Genes Dev.* 8:970-980). The alpha helical regions are usually 15-16 amino acids long with hydrophobic residues at every third and fourth position, and each helix contains several conserved residues (Murre C et al. (1989) *Cell*, 56:777-783; Benezra R. et al. (1990) *Cell*, 61:49-59).

The HLH protein family is subdivided into two major groups: the so-called "bHLH" and "non basic HLH" subfamilies. Proteins of the bHLH family contain a conserved highly basic
25 region immediately N-terminal to the first helix (known as bHLH structure), and mutagenesis experiments on MyoD protein confirm that this region is responsible for sequence-specific binding to the "E-box", a consensus DNA motif for bHLH proteins (Davis RL. et al. (1990) *Cell*, 60: 733-746). A dimeric bHLH protein (either homodimeric or heterodimeric but in which both subunits contains a basic region) are able to bind to DNA. In general, the bHLH proteins fall into two
30 categories: Class A consists of proteins that are ubiquitously expressed, including mammalian E12/E47 and fly *da* whereas the class B consists of proteins that are expressed in a more tissue-specific manner, including mammalian MyoD and fly AC-S. In most cases, the tissue-specific bHLH proteins preferentially heterodimerize with ubiquitous partners.

The non basic HLH subfamily contains proteins lacking a basic region unable to bind to
35 DNA but that could form homo- or heterodimers through their HLH motif. Indeed, heterodimeric complexes between non basic HLH and bHLH proteins fail to bind to DNA and negatively modulate the bHLH proteins-mediated transcription activation. This phenomenon was first

- demonstrated in a MyoD/Id regulation model (Benezra R. et al. (1990) Cell, 61:49-59). The MyoD gene product is able to activate previously silent muscle-specific genes when introduced into a large variety of differentiated cell types. MyoD proteins form either homodimers or heterodimers with other bHLH proteins such as E12 or E47, and bind to E-box consensus motif to activate
- 5 myogenesis. The Id gene, conserved from batracians to mammals (Wilson R et al. (1995) Mech.Dev. 49:211-222; Sawai S et al. (1997) Mech.Dev. 65:175-185; Norton JD et al. (1998) trends in Cell Biology 8:58-65), lacks a basic region adjacent to its HLH motif but is able to specifically dimerize with either MyoD, E12 or E14 and has been shown to subsequently attenuate the heterodimer's ability to bind DNA. Additionally, overexpression of Id inhibits MyoD-
- 10 dependent gene activation in in vivo transfection experiments. Id proteins may function either to repress directly the activity of tissue-restricted bHLH proteins by rendering them non-functional or, more likely, to sequester the ubiquitous bHLH proteins and preventing them from forming active heterodimers with the tissue-restricted bHLH (Review by Norton JD et al. (1998) trends in Cell biology 8:58-65).
- 15 The possibility that the Id protein behaves as a dominant-negative regulator to repress MyoD protein activity through the formation of nonfunctional heterodimeric complexes is considerably strengthened by the following findings in Drosophila. In Drosophila, the development of peripheral nervous system is positively regulated by the two structurally related bHLH proteins, AS-C and daughterless (da), since loss of either activity results in loss of sensory organ
- 20 development. The extramacrochaetae Emc product belonging to the non basic HLH subfamily was shown to antagonize the activity of AS-C and da. through the formation of nonfunctional heterodimers with the bHLH proteins (Hillary M et al. (1990) Cell, 61:27-38; Garrell J et al. (1990) Cell 61,39-48).
- Human Id genes including human Id1, Id2, Id3 and Id4 have been identified and localized
- 25 (Review by Norton JD et al. (1998) trends in Cell Biology 8:58-65). The bHLH proteins and Id proteins are thought to be involved in the regulation of apoptosis. Differentiation and development of T- and B-lymphocytes in immune system are positively regulated by the combination of ubiquitous E proteins and lymphocyte-restricted bHLH proteins. Disruption in gene expression from either class results in severe perturbation of T- and B-lymphocyte development (Bain G et al. (1997) Mol Cell Biol 17:4782-4791; Zhuang et al. (1996) Mol Cell Biol 16:2898-2905). Cell-
- 30 arrested T thymocytes undergo a massive apoptosis when Id1 gene is overexpressed (Kim D (1999) Mol Cell Biol 19(12):8240-53). Overexpression of Id1 gene product also results in apoptosis in neonatal and adult cardiac myocytes in culture (Tanaka K et al. (1998) J Biol Chem 273(40) 25922-25928).
- 35 Id1 and Id3 proteins are also required to support angiogenesis. Quiescent adult endothelial cells express minimal level of the Id proteins, whereas Id expression is upregulated in angiogenic endothelial cells. Partial loss of these proteins in Id1^{-/-}Id3^{-/-} double knockout mice impairs

angiogenesis, resulting in the resistance to tumour growth (Lyden D et al. (1999) Nature 401:670-677). In addition, a significant overexpression of mRNA and protein levels of Id1, Id2 and Id3 has been found in patients with pancreatic cancer (Maruyama H et al. Am J Pathol (1999) 155(3):815-822) A correlation of Id1 gene upregulation and aggressive phenotype of human breast cancer cells has also been reported (Lin CQ et al. (2000) Cancer Res 60(5):1332-40).

Thus, identification and cloning of members of the HLH family, and especially of the non basic HLH subfamily, is necessary to enrich our knowledge about the biological importance of the HLH transcription factors network and further more to provide insights and tools in disorders linked to dysregulation of the HLH-mediated transcription.

10 It is believed that the protein of SEQ ID NO: 394 or part thereof plays a role in the regulation of transcription activation, probably as a member of the HLH family, preferably of the non basic HLH subfamily. More particularly, the protein of the invention is thought to be able to antagonize the activity of members of the bHLH family through the formation of heterodimers. Preferred polypeptides of the invention are polypeptides comprising the amino acids of SEQ ID
15 NO: 394 from positions 13 to 28, from positions 53 to 68, and from positions 13 to 68. Other preferred polypeptides of the invention are fragments of SEQ ID NO: 394 having any of the biological activity described herein.

The dimerization ability of the protein of the invention or part thereof which is characteristic of the HLH family may be assayed using any of the assays known to those skilled in
20 the art. For example, interacting protein partners, especially members of the bHLH subfamily, may be identified using screening of cDNA expression libraries as described for the identification of some HLH transcription factors such as E12 and E47 (Murre C et al. (1989) Cell 56:777-783), Max (a Myc binding factor) (Elizabeth M et al. (1991) Science 251:1217) as well as Id (Benezra C et al. (1990) Cell 61:49-59). Alternatively, the helix-loop-helix motif in the protein of the invention
25 could be used by those skilled in art as a "bait protein" in a well established yeast double hybridization system to identify its interacting protein partners in vivo from cDNA library derived from different tissues or cell types of a given organism. Alternatively, the protein of the invention or part thereof could be used by those skilled in art in mammalian cell transfection experiments. When fused to a suitable peptide tag such as [His]₆ tag in a protein expression vector and introduced
30 into culture cells, this expressed fusion protein can be immunoprecipitated with its potential interacting proteins by using anti-tag peptide antibody. This method could be chosen either to identify the associated partner or to confirm the results obtained by other methods such as those just mentioned.

An object of the invention relates to compositions and methods using the protein of the
35 invention or part thereof to dysregulate gene transcription, preferably transcription mediated by HLH regulators either in vitro or in vivo, through overexpression of the protein of the invention using any means known to those skilled in the art.

The protein of the invention or part thereof could be used to induce apoptosis of specific cell-type under either physiological or pathological conditions. In a preferred embodiment, the apoptosis active polypeptide is added to an in vitro culture of mammalian cells in an amount effective to induce apoptosis. In another preferred embodiment, the apoptosis active polypeptide is expressed under the control of a promoter which may be activated under precise conditions. In particular, such conditional expression of an apoptosis-active polypeptide upon demand may be very useful to get rid of cells that have become unwanted, for example in applications where such cells have been used in a cellular therapy goal and have become useless. Another example of application is the case of expression under the control of a promoter that becomes active after infection by a given microorganism, thus resulting in the death of the infected cells only. Furthermore, the protein of the invention or part thereof may be useful in the diagnosis, the treatment and/or the prevention of disorders in which apoptosis is beneficial, including but not limited to disorders linked to abnormal cellular proliferation such as those described below.

In another embodiment, the protein of the invention or part thereof can be used to diagnose, treat and/or prevent disorders linked to overexpression of HLH proteins, such as cancer and other disorders relating to abnormal cellular differentiation, proliferation, or degeneration, including hyperaldosteronism, hypocortisolism (Addison's disease), hyperthyroidism (Grave's disease), hypothyroidism, colorectal polyps, gastritis, gastric and duodenal ulcers, ulcerative colitis, and Crohn's disease, neurodegenerative disorders such as Parkinson's or Alzheimer's diseases using any methods and/or techniques described herein. In addition, the protein of the invention or part thereof may be used to evaluate the disease progression and the clinical treatment efficiency. The protein of the invention or part thereof could also be used a molecular target for anti-angiogenesis drug design. Inhibition of protein expression could be achieved by many means known to those skilled in the art including those described in the present application. For example, an antisense nucleotide or triple helix strategy could be developed to block the protein synthesis. Alternatively, the expressed protein of the invention might be neutralized by using specific monoclonal antibody using techniques known to those skilled in the art including those described in Peverali FA et al (1994) EMBO J. 13:4291-4301; Barone MV et al. (1994) Proc.Natl.Acad.Sci.USA 91:4985-4988; and Haza ET et al. (1994) J.Biol.Chem. 269:2139-2145.

Protein of SEQ ID NO: 466 (internal designation 184-4-2-0-D3-CS)

The protein of SEQ ID NO: 466 overexpressed in liver and encoded by the cDNA of SEQ ID NO: 225 displays a Zinc finger motif of RING type (C3HC4) (Pfam signature from positions 41 to 81, Prosite signature from positions 56 to 65) and a B-box zinc finger motif (pfam signature from positions 110 to 153). In addition, the protein of the invention is predicted to have a nuclear localization.

It is believed that the protein of SEQ ID NO: 466 or part thereof is a zinc binding protein, preferably able to bind nucleic acids, more preferably a transcription factor. Preferred polypeptides of the invention are polypeptides comprising the amino acids of SEQ ID NO: 466 from positions 41 to 81 (Ring Zinc finger protein), and from 110 to 153 (B-Box domain). Other preferred
5 polypeptides of the invention are fragments of SEQ ID NO: 466 having any of the biological activity described herein.

Protein of SEQ ID NO: 267 (internal designation 116-111-1-0-H9-CS)

The protein of SEQ ID NO: 267 encoded by the extended cDNA SEQ ID NO: 26 exhibits an Emotif zinc finger domain, C2H2 type, from positions 185 to 202, and is thought to be localized
10 in the nucleus.

It is believed that the protein of SEQ ID NO: 267 or part thereof is a zinc binding protein, preferably able to bind nucleic acids, more preferably a transcription factor. Preferred polypeptides of the invention are polypeptides comprising the amino acids of SEQ ID NO: 267 from positions 185 to 202. Other preferred polypeptides of the invention are fragments of SEQ ID NO: 267 having
15 any of the biological activity described herein.

Protein of SEQ ID NO: 277 (internal designation 160-103-1-0-F11-CS)

The protein of SEQ ID NO: 277 encoded by the extended cDNA SEQ ID NO: 36 exhibits a pfam DHHC zinc finger domain from positions 140 to 204.

It is believed that the protein of SEQ ID NO: 277 or part thereof is a zinc binding protein,
20 preferably able to bind nucleic acids, more preferably a transcription factor. Preferred polypeptides of the invention are polypeptides comprising the residues of SEQ ID NO: 277 from positions 140 to 204. Other preferred polypeptides of the invention are fragments of SEQ ID NO: 277 having any of the biological activity described herein.

Protein of SEQ ID NO: 272 (internal designation 145-25-3-0-B4-CS)

25 The protein of SEQ ID NO: 272 encoded by the extended cDNA SEQ ID NO: 31 shows homology with numerous zinc binding proteins. In addition, the protein of the invention exhibits the pfam RING zinc finger signature from positions 87 to 129. The protein of SEQ ID NO: 272 has a variant, i.e. the protein of SEQ ID NO: 273 encoded by the extended cDNA SEQ ID NO: 32 and thought to have the same function and utilities.

30 It is believed that the protein of SEQ ID NO: 272 or part thereof is a zinc binding protein, preferably able to bind nucleic acids or proteins, more preferably a transcription factor. Preferred polypeptides of the invention are polypeptides comprising the amino acids of SEQ ID NO: 272 from positions 87 to 129. Other preferred polypeptides of the invention are fragments of SEQ ID NO: 272 having any of the biological activity described herein.

Hydrolases and inhibitors

The invention relates to compositions and methods using proteins of the invention having a hydrolytic activity, herein referred to as HYP, such as the ones described in this section and those containing an hydrolytic domain as shown on Table VI, or parts thereof, preferably fragments comprising an hydrolytic domain, or derivative thereof.

The invention relates to methods and compositions using HYP or a fragment thereof to hydrolyze one or several substrates, alone or in combination with other substances.. For example, the protein of the invention or part thereof is added to a sample containing the substrate(s) in conditions allowing hydrolysis, and allowed to catalyze the hydrolysis of the substrate(s). Hydrolyzed substrates are then detected using standard methods known to those skilled in the arts. The protein of the invention or part thereof can also be added to samples as a "cocktail" with other hydrolytic enzymes, such as other peptidases, for example to decontaminate surgical instruments using methods described in US patent 5,489,531. The advantage of using a cocktail of hydrolytic enzymes is that one is able to hydrolyze a wide range of substrates without necessarily knowing the specificity of each enzyme. Using a cocktail of hydrolytic enzymes also protects a sample from a wide range of future unknown contaminants from a vast number of sources. Alternatively, HYP or part thereof may be bound to a chromatographic support, either alone or in combination with other hydrolytic enzymes, using techniques well known to those skilled in the art, to form an affinity column to remove the substrate. Immobilization facilitates removal of the enzyme from the batch of product and subsequent reuse of the enzyme.

Immobilization of the enzyme or part thereof can be accomplished, for example, by adding a cellulose-binding domain to the protein through the modification of the DNA sequence coding for the protein or part thereof. One of skill in the art will understand that other methods of immobilization could also be used and are described in the available literature. Alternatively, the same methods may be used to identify new substrates.

In another embodiment, HYP or part thereof may be used to identify or quantify the amount of a given substrate in a biological sample. In a preferred embodiment, HYP or part thereof is catalytically inactivated, i.e. capable of binding but not hydrolyzing a given substrate, using any of the methods known to those skilled in the art including those which produce a mutant enzyme, a recombinant-enzyme, or a chemically inactivated enzyme. The catalytically inactive protein of the invention is then incubated with an aliquot of a biological sample under conditions suitable for binding of the inactive enzyme to the substrate. Then, the bound enzyme is detected to assess the presence or amount of the eubacteria in the biological sample. In another preferred embodiment, HYP or part thereof is used in assays and diagnostic kits for the identification and quantification of substrates in a biological sample. These assays can be based for example, on standard enzyme-linked immunosorbent assays (ELISA) or any other technique known to those skill in the art. In addition, HYP or part thereof may be used to identify, e.g. using screens based on standard assays

such as those described above, inhibitors of the enzyme for mechanistic and clinical applications. Such inhibitors may then be used to identify or quantify HYP in a sample, and to diagnose, treat or prevent any of the disorders where the protein's activity is undesirable and/or deleterious.

Protein of SEQ ID NO:400 (internal designation 160-54-1-0-F7-CS)

5 The protein of SEQ ID NO:400, encoded by the cDNA of SEQ ID NO:159, exhibits two putative transmembrane domains encompassing amino-acids 50-70 and 127-147 as predicted by the software TopPred II (Claros and von Heijne, *CABIOS applic. Notes*, 10 :685-686 (1994)). It also displays the Prosite carboxypeptidase zinc-binding region signature PS00133 at positions 117-127. It is predicted by the psort software (see Nakai K and Horton P, *Trends Biochem Sci.* 1999 Jan;24(1):34-6) to localize to the nucleus with a high probability (73.9%). Finally it is specifically expressed in fetal brain and shows no homology to previously known proteins.

 Carboxypeptidase enzymes hydrolyze the terminal amino acid of a protein or peptide. A novel family of carboxypeptidases, localized in the nucleus and with a carboxypeptidase-dependant transcriptional activity, has emerged only recently. Its first member, AEBP1, was previously
15 identified as a 3T3 preadipocyte factor implicated in the repression of the aP2 gene expression. AEBP1 stands for "AE-1 Binding Protein," where AE-1 is a regulatory element of the adipose P2 gene (aP2), a gene involved in triglyceride metabolism and activated in adipocytes. Its own expression is abolished during adipocyte differentiation (He GP et al., *Nature* 378:92-96(1995)). AEBP1 was subsequently shown to play a similar role in the differentiation of osteoblastic cell lines
20 (Ohno I et al., *Biochem Biophys Res Commun.* 1996 Nov 12;228(2):411-4) and vascular smooth muscle cells (Layne MD et al., *J. Biol. Chem.* 273:15654-15660(1998)). It was proposed that AEBP1 acts as a negative transcription factor by cleaving proteins involved in transcription, a new feature in transcription regulation. Recent evidence further suggests that its transcriptional activity is itself attenuated by binding to G-proteins subunits (Park JG et al., *EMBO J.* 1999 Jul
25 15;18(14):4004-12) and stimulated by DNA binding (Muisse AM and Ro HS, *Biochem J.* 1999 Oct 15;343 Pt 2:341-5).

 It is believed that the protein of SEQ ID NO:400 plays a role in cell signaling, nuclear transcriptional activity and in the differentiation of several cell types, especially those found in the developing brain (including but not limited to neurons). Preferred polypeptides of the invention are
30 polypeptides having any of the biological activities described herein.

 One embodiment of the present invention relates to compositions and methods using the protein of the invention or part thereof as a marker for specific cell compartments (especially the nucleus) and/or tissue types (especially fetal brain). For example, the protein of the invention or part thereof may be used to generate specific antibodies which would in turn allow the visualization
35 of nuclear structures by methods well-known to those of skill in the art. In a similar fashion, antibodies raised against the protein of the invention may be used to identify particular

developmental stages (fetal for instance) and/or given tissue types (brain for instance), as the protein of the invention is specifically expressed in brain tissues at a fetal stage. Antibodies and antiserum can also be used to inhibit undesirable carboxypeptidase activities in *in vitro* experiments and cell cultures, as well as in biological samples and *in vivo*. Alternatively, quantitative analysis or
5 detection of the protein of the invention, or of nucleic acids encoding the protein, can be carried out by any other technique known to those skilled in the art.

In another embodiment, the protein of the invention may be used to target heterologous compounds (polypeptides or polynucleotides) to the developing brain and/or the cell nucleus. For instance, a chimeric protein composed of the protein of the invention recombinantly or chemically
10 fused to a protein or polynucleotide of therapeutic interest would allow the delivery of the therapeutic protein/polynucleotide specifically to the above-mentioned cellular/tissue targets (nucleus, fetal brain).

In another embodiment, the present invention relates to methods and compositions using the protein of the invention or a fragment thereof to hydrolyze one or several substrates, alone or in
15 combination with other substances. The ability of the present protein to hydrolyze any particular substrate can easily be determined by carrying out a hydrolysis reaction using standard assay techniques such as the ones described by Slusher et al. (Slusher et al. – Prostate – 2000, 44(1): 55-60) or any other technique well known to those skilled in the art. Potential substrates are any substance containing a peptide bond, more specifically a C-terminal peptide bond. Such substances
20 include, but are not limited to, polypeptides, folic acid and its analogues (e.g. methotrexate). For example, the protein of the invention or part thereof is added to a sample containing the substrate(s) in conditions allowing hydrolysis, and allowed to catalyze the hydrolysis of the substrate(s). Hydrolyzed substrates are then detected using standard methods known to those skilled in the art.

In a preferred embodiment, the protein of the invention or part thereof may be used to
25 modulate cellular transcriptional activity, thereby modulating cellular differentiation. Specifically, as nuclear carboxypeptidases play a role in inhibiting transcription associated with differentiation, then an increase in the activity or expression of the protein can be used to inhibit differentiation. The ability to inhibit differentiation has a number of uses, for example during the cultivation of undifferentiated pluripotent cells to maintain the cultured cells in an undifferentiated state until the
30 need for a given cell type arises (in cases of grafts for instance). The level of the protein activity or expression can be increased in any of a number of ways, including by introducing a polynucleotide encoding the protein into cells, by administering the protein itself to cells, or by administering to cells a compound that increases protein activity or expression. Alternatively, the protein of the invention can be inhibited, thereby enhancing cellular differentiation. The ability to promote
35 differentiation has many uses, including in the treatment or prevention of cancer, as cancer cells are often in a relatively undifferentiated state, and cellular differentiation typically accompanies by growth arrest.

In another embodiment, the protein of the invention or part thereof may be used to diagnose, treat and/or prevent disorders where the presence of substrates, for example excess proteins or peptides, is undesirable or deleterious. Such disorders include but are not limited to, cancer, neurodegenerative disorders such as Parkinson's and Alzheimer's diseases, and diabetes. In another embodiment, the protein of the invention or part thereof may be used to identify or quantify the amount of a given substrate (e.g. a peptide, folic acid, or methotrexate) in a biological sample. In a preferred embodiment, the protein of the invention or part thereof is used in assays and diagnostic kits for the identification and quantification of substrates in a biological sample.

In a most preferred embodiment, the protein of the invention or part thereof can be used in cancer chemotherapies in rescue therapy following toxic high dose methotrexate regimes. Many carboxypeptidases can cleave the C-terminal glutamate moiety from folic acid and its analogues, such as methotrexate. The key role of reduced folates as coenzymes in many biological pathways including those leading to DNA synthesis via the pyrimidines and purines, has made folic acid a target molecule for chemotherapy. Tumor cells grow rapidly and have a high rate of nucleic acid synthesis. Depletion of folic acid has cytotoxic effects, primarily in replicating tissues, and can inhibit growth of tumors with high folic acid requirements. Many carboxypeptidases can directly deplete folate by hydrolytic removal of its glutamate moiety. In cancer chemotherapy, methotrexate (4-amino-N¹⁰-methyl-pteroyl-glutamate) is commonly used to deplete the pool of reduced folates by inhibiting dihydrofolate reductase (DHFR), which catalyses the reduction of folates into biologically active tetrahydrofolate form, essential in the biosynthesis of all folate coenzymes. Thus, the protein of the invention or part thereof could be used in rescue therapy following toxic high-dose regimes such as described by Widemann et al. (Widemann B. et al. – Proc. Am. Assoc. Cancer Res. – 1995, 36, p232) and Chabner et al. (Chabner B. et al. – Nature – 1972, 239, p395-397), which disclosures are hereby incorporated by reference in their entirety. The basis of this strategy is that hydrolysis of methotrexate produces 4-amino-N¹⁰-methyl-pterotate that is about 100 times less active as an inhibitor of DHFR.

In another preferred embodiment, the protein of the invention or part thereof can be used in an enzyme/prodrug strategy to treat a number of pathologies, especially those treated with drugs associated with severe side effects, including, but not limited to, autoimmune diseases and chronic inflammatory diseases such as rheumatoid arthritis, and cancer chemotherapy. These side effects can be mainly explained by the fact that the in vivo selectivity of the drugs used is too low (for example, the inadequate selectivity between tumor and normal cells of most anticancer drugs is well known and their toxicity to normal tissues is dose limiting). In the first phase of one example of such a protocol, a conjugate of the protein of the invention or part thereof and an antibody to a tissue specific antigen (for example, tumor specific antigens in the case of cancer chemotherapy) is administered. After a delay to allow residual enzyme conjugate to be cleared from the blood, a relatively non-toxic compound is administered to the patient. This non-toxic compound is a

substrate of the protein of the invention, and is converted by the protein into a substantially more toxic compound. Thus, because of the previous, targeted administration of the protein of the invention, when the non-toxic compound is administered, the toxic compound is only produced in the vicinity of the cells targeted by the fusion protein. This two-phase approach has been termed
5 antibody-directed enzyme-prodrug therapy (ADEPT), this approach is reviewed by Melton et al. (Melton R. et al. – J. Natl. Cancer Inst. – 1996, 88, p153-165). Alternatively the first phase can be replaced by a gene therapy approach resulting in the de novo synthesis of the protein of the invention or part thereof by cells from the targeted tissue, this has been termed gene-dependent enzyme/prodrug therapy (GDEPT). Another advantage of these 2 approaches (ADEPT and
10 GDEPT) is that a single enzyme molecule is capable of activating many prodrug molecules.

Protein of Seq Id No: 242 (internal designation 119-003-4-0-C2-CS)

The protein of SEQ ID No: 242, encoded by the cDNA of SEQ ID No: 1, is homologous to proteins of the M20 metallopeptidases family (EC 3.4.17.X). The protein of the invention is over-expressed in the spinal cord and the brain.

15 The M20 metallopeptidase family of proteins are all peptidases (i.e. enzymes able to hydrolyze peptide bonds) furthermore they are all exopeptidases, which means that they can hydrolyze the terminal amino acid of a protein or peptide. Members of the M20 peptidase family are glutamate carboxypeptidases, which are capable of releasing the C-terminal glutamate residue, by hydrolysis, from a wide range of N-acyl groups, including peptidyl, aminoacyl, benzoyl,
20 benzyloxycarbonyl, folyl, and pteroyl groups, and physiologically are involved in the catabolism of proteins. M20 carboxypeptidases are either monomeric or homodimeric (i.e. 2 identical proteins assembled to form the enzyme). In order to be active, metallopeptidases must be associated with a metallic cofactor (either Zinc or Cobalt depending on the enzyme). The most studied carboxypeptidase of the M20 family is carboxypeptidase G2 (CPG2) (EC 3.4.17.11), a bacterial
25 enzyme from *Pseudomonas* sp. (strain RS-16). CPG2 is a dimeric Zinc carboxypeptidase that cleaves the C-terminal glutamate moiety from a number of molecules.

The protein of SEQ ID No: 242 includes the pfam signature for M20 peptidase (position 107 to 451). The protein of SEQ ID No: 242 also includes a number of amino acids that are conserved throughout the M20 protease family especially those that interact with the metal cofactor.
30 Preferred polypeptides of the invention are polypeptides of SEQ ID No: 242 that include the highly conserved amino acids: 133, 135, 149, 163, 200, 201 and/or 262, which are present in over 80% of the members of the M20 peptidase family, and/or amino acids 139, 157, 162, 16, 367 and/or 377, which are present in over 60% of the members of the M20 peptidase family. Of particular interest are amino acids 133, 166, 201 and 262, which by homology are probably involved in the interaction
35 with the metal cofactors. Thus it is believed that the protein of SEQ ID No: 242 or part thereof is a peptidase, preferably a carboxypeptidase, more preferably a metallocarboxypeptidase of the M20

family. Other preferred polypeptides of the invention are any fragments of SEQ ID No: 242 having any of the biological activities described herein.

Determination of carboxypeptidase activity on specific substrates can easily be obtained by carrying out the hydrolysis using standard assay techniques such as the ones described by Slusher et al. (Slusher et al. – Prostate – 2000, 44(1): 55-60) or any other technique well known to those skilled in the art. Potential substrates are any substance containing a peptide bond, more especially C-terminal peptide bonds, and even more specifically, C-terminal glutamate. Such substances include but are not limited to peptides, folic acid and its analogues (e.g. methotrexate).

In an embodiment the protein of the invention or part thereof could be used to develop assay tools to identify brain and spinal cord tissue since the protein of the invention is overexpressed in these tissues.

In still another embodiment, the protein of the invention or part thereof may be used to diagnose, treat and/or prevent disorders where the presence of substrates, for example excess proteins, is undesirable or deleterious. Such disorders include but are not limited to, cancer, neurodegenerative disorders such as Parkinson's and Alzheimer's diseases, and diabetes. In a most preferred embodiment, the protein of the invention or part thereof can be used in cancer chemotherapies in rescue therapy following toxic high dose methotrexate regimes. Enzymes of the M20 peptidase family can cleave the C-terminal glutamate moiety from folic acid and its analogues, such as methotrexate. The key role of reduced folates as coenzymes in many biological pathways including those leading to DNA synthesis via the pyrimidines and purines, has made folic acid a target molecule for chemotherapy. Tumor cells grow rapidly and have a high rate of nucleic acid synthesis. Depletion of folic acid has cytotoxic effects, primarily in replicating tissues, and can inhibit growth of tumors with high folic acid requirements. Enzymes of the M20 peptidase family can directly deplete folate by hydrolytic removal of its glutamate moiety. In cancer chemotherapy, methotrexate (4-amino-N¹⁰-methyl-pteroyl-glutamate) is commonly used to deplete the pool of reduced folates by inhibiting dihydrofolate reductase (DHFR), which catalyses the reduction of folates into biologically active tetrahydrofolate form, essential in the biosynthesis of all folate coenzymes. Thus the protein of the invention or part thereof could be used in rescue therapy following toxic high-dose regimes such as described by Widemann et al. (Widemann B. et al. – Proc. Am. Assoc. Cancer Res. – 1995, 36, p232) and Chabner et al. (Chabner B. et al. – Nature – 1972, 239, p395-397), which disclosures are hereby incorporated by reference in their entirety. The basis of this strategy is that hydrolysis of methotrexate produces 4-amino-N¹⁰-methyl-pterate that is about 100 times less active as an inhibitor of DHFR.

In another preferred embodiment, the protein of the invention or part thereof can be used in an enzyme/prodrug strategy to treat a number of pathologies, especially those treated with drugs associated with severe side effects, including, but not limited to, autoimmune diseases and chronic inflammatory diseases such as rheumatoid arthritis, and cancer chemotherapy. These side effects

can be mainly explained by the fact that the in vivo selectivity of the drugs used is too low (for example, the inadequate selectivity between tumor and normal cells of most anticancer drugs is well known and their toxicity to normal tissues is dose limiting). In the first phase of one example of such a protocol, a conjugate of the protein of the invention or part thereof and an antibody to a tissue specific antigen (for example, tumor specific antigens in the case of cancer chemotherapy) is administered. After a delay to allow residual enzyme conjugate to be cleared from the blood, a relatively non-toxic compound is administered to the patient. This non-toxic compound is a substrate of the protein of the invention, and is converted by the protein into a substantially more toxic compound. Thus, because of the previous, targeted administration of the protein of the invention, when the non-toxic compound is administered, the toxic compound is only produced in the vicinity of the cells targeted by the fusion protein. This two-phase approach has been termed antibody-directed enzyme-prodrug therapy (ADEPT), this approach is reviewed by Melton et al. (Melton R. et al. – J. Natl. Cancer Inst. – 1996, 88, p153-165). Alternatively the first phase can be replaced by a gene therapy approach resulting in the de novo synthesis of the protein of the invention or part thereof by cells from the targeted tissue, this has been termed gene-dependent enzyme/prodrug therapy (GDEPT). Another advantage of these 2 approaches (ADEPT and GDEPT) is that a single enzyme molecule is capable of activating many prodrug molecules.

Protein of SEQ ID NO: 401 (internal designation 160-88-3-0-A8-CS.corr)

The protein of SEQ ID NO : 401 encoded by the cDNA SEQ ID NO: 160 is a splicing variant of the hypothetical human palmitoyl-protein thioesterase-2 (PPT2) (E.C. 3.1.2.22) (Genbank accession number AF020543), which is well conserved among eukaryotes (*C. elegans* and rodents) and exhibits homology with the palmitoyl protein thioesterase-1 (PPT1) (Genbank accession number L42809). The product of the cDNA SEQ ID NO: 160 is shorter than the human PPT2 (280 versus 308 amino acids respectively) with a gap located between the positions 174 and 203 of the protein PPT2. The protein of SEQ ID NO : 401 has a variant, the protein of SEQ ID NO: 402 encoded by the cDNA of SEQ ID NO: 161, thought to have the same functions and utilities.

PPT1 (E.C. 3.1.2.22) is a well-described protein, widely conserved among the murine, rat, bovine and human species (Swissprot accession number P50897). It is a lysosomal enzyme that functions in the removal of fatty acids from modified cysteine residues in proteins undergoing degradation (Hofmann S.L. *et al*, *Neuropediatrics*, **28**: 27-30 (1997)). For example, PPT1 catalyses the deacylation H-ras and the alpha subunits of heterodimeric G proteins *in vitro* (Camp L.A., *J. Biol. Chem.*, **268**: 22566-22574 (1993) and **269**: 23212-23219 (1994)). Deacylation by PPT1 may be a prerequisite for complete digestion of the modified polypeptides. In fact there is evidence that palmitoylation leads to increased protection against proteolytic digestion. Both the salivary mucus glycoprotein (Slomiany B. L., *Biochem. Biophys. Res. Commun.*, **151**: 1046-1053 (1988),) and chemically acylated bee venom phospholipase A2 (Diaz, R.E., *Biochem. Biophys. Acta*, **830**: 52-58

(1985)) are more resistant to treatment with proteinases than their deacylated forms. Mutations in PPT1 enzyme were shown to underlie the hereditary neurodegenerative disorder, infantile neuronal ceroid lipofuscinosis (Vesa *et al.*, *Nature*, 376: 584-587 (1995)).

Recently, Soyombo and Hofmann (*J.Biol.Chem*, 272: 27456-27463, (1997)) described a second lysosomal thioesterase, PPT2, that shares 20% identity with PPT1. The PPT2 enzyme presumably also plays a role in lysosomal thioester catabolism but has a substrate specificity distinct from that of PPT1. While little is known about the substrate specificity of PPT2, the enzyme is highly active against palmitoylated model substrates such as palmitoyl CoA. PPT2 did not hydrolyse the acyl-cysteine bond of the protein substrates routinely used to assay PPT1 such as H-Ras and albumin. This finding suggest that although both enzymes possess intrinsic palmitoyl thioesterase activity, the "leaving group" recognized by the enzymes may differ. One possibility is that PPT2 recognizes palmitoylated protein substrates but that these substrates differ from those recognized by PPT1. A second possibility is that PPT2 recognizes a novel lipid thioester substrate that is not derived from acylated proteins. Aguado *et al.* (*Biochem J.*, 341:679-689, (1999)) demonstrated that PPT2 is an acyl thioesterase. However they cannot distinguish between esterase (thioesterase) and lipase activity. PPT2 shows very high S-thioesterase activity towards the acyl chains $C_{14:0} > C_{16:0}$, moderate activity towards the acyl chains $C_{14:1} > C_{20:4} \approx C_{16:1} \approx C_{18:0} \approx C_{12:0} > C_{18:2} \approx C_{18:3} > C_{22:1} \approx C_{18:1} \approx C_{20:0}$, low activity towards the acyl chains $C_{10:0}$ and $C_{22:0}$, and no activity towards the acyl chain $C_{24:0}$, $C_{8:0}$, $C_{6:0}$, $C_{4:0}$ and $C_{2:0}$. PPT2 has a broader range of action than PPT1, although both have a preference for long acyl chains (more than 12 or 14 carbons) over shorter acyl chains (less than 12 carbons). Aguado *et al.* (supra) also presented a detailed characterization of PPT2 gene product. The putative 302-residue PPT2 and the protein of the invention contains a hydrophobic leader peptide at the N-terminus (signal peptide with a cleavage site predicted at position 34 of the protein of the invention) suggesting that they are secretory glycoproteins. Both proteins exhibit two motifs located at the N-terminus from positions 108 to 121. One motif is common to triglycerides lipases (from position 110 to 121) and the other one to eukaryotic thiol (Cys) proteases (from positions 108 to 121). Triglyceride lipases are lipolytic enzymes that hydrolyse the ester bond of triglycerides. The most conserved region in all these proteins is centered on a serine residue located in a conserved Gly-Xaa-Ser-Xaa-Gly motif. The PPT2 protein and the protein of the invention contain a cysteine residue (position 115) instead of the first glycine residue in the motif but other lipases with one mismatch in either of the consensus have been described (Blow D., *Nature*, 343: 694-695 (1990)). In the same region as the lipase motif, PPT2 and the protein of the invention contains a motif common to the active site of eukaryotic thiol (Cys) protease but with a leucine residue (position 113) instead of the glycine at the position 5 of the pattern. In addition, the amino acid sequence of the putative PPT2 shows, at the C-terminus, from positions 171 to 186, a motif common to growth factor and cytokine receptors family, which is not present in the protein of the invention.

Aguado *et al.* (supra) have found that PPT2 is expressed in cells of the immune system as an approximatively 42 kDa protein in cells extracts and supernatants and is transcribed as at least five different transcripts. The PPT2 gene is located in the class III region of the human MHC which contains several genes encoding proteins with potential roles in the immune system and in inflammation. In addition, Aguado *et al.* (supra) showed that very large amounts of PPT2 are secreted. However this is not in disagreement with an intracellular activity because the secreted protein could be internalized into the cell through a receptor and act on target located in an intracellular organelle. This mechanism has been described for the secreted PPT1, which can be internalized into the cell by mannose-6-phosphate receptor to act in the lysosome (Verkruyse and Hofmann, *J.Biol.Chem.*, **271**: 15831-15836, (1996)), and Soyombo and Hofmann (*J.Biol.Chem.*, **272**: 27456-27463, (1997)) reported that PPT2 binds to mannose-6 phosphate receptor.

Palmitoylation refers to posttranslational modification of proteins in which the most common fatty acids of the cell (i.e. palmitic, stearic and oleic acids) are attached to the side chain of cysteine residues via high-energy thioester linkages (Bizzozero, O.A. *et al*, *Neurochem.Res.*, **19**: 923-933 (1994); Casey P.J., *Science*, **268**: 221-225 (1995)). At present a large number of proteins of diverse origin, structure and function are known to be modified with these fatty acids that attach them to inner surface of the plasma membrane, where they can function optimally (Casey P.J., *Science*, **268**: 221-225 (1995)). Being anchored to membranes is a process necessary for the diverse cellular functions of these modified proteins, including signal transduction, vesicle transport and maintenance of the cytoarchitecture. Almost every tissue and subcellular organelle contains characteristic set of palmitoylated proteins.

The protein of the invention is overexpressed in brain. In recent years a considerable number of functionally relevant nervous system proteins including ion channels, neurotransmitter receptors, signal transduction components and cell-adhesion molecules have been found to be palmitoylated. Although the nervous system is not an exception to this rule, both the number of modified protein in this tissue and the dynamic nature of protein palmitoylation suggest that this modification is critical for regulating important biological processes and that the addition or removal of the fatty acid serves to regulate the activity of these proteins rather than to define their function.

It is believed that the protein of SEQ ID NO: 401 or part thereof is an hydrolase, preferably acting on ester bonds, more preferably a thiolester hydrolase, even more preferably an acyl-thioesterase which, as such, plays a role in fatty acid metabolism, in cellular vesicle transport and maintenance of the cytoarchitecture, in cellular proteolysis, endocytosis, signal transduction, lysosomal storage, cell proliferation and differentiation, immune and inflammatory response. The enzyme's substrates are compounds preferably containing an ester bond, preferably a thiol ester bond, more preferably an acyl thioester bond. Preferred polypeptides of the invention are polypeptides comprising the amino acids of SEQ ID NO: 401 from positions 108 to 121, and 110 to

121. Other preferred polypeptides of the invention are fragments of SEQ ID NO: 401 having any of the biological activities described herein. The hydrolytic activity of the protein of the invention or part thereof may be assayed using any of the assays known to those skilled in the art including those described in Smith *et al.*, *Biochem.J.*, **212**: 155 (1983), Spencer *et al.*, *J.Biol.Chem.*, **253**: 5922
5 (1978) and Aguado *et al.* (supra) or in US patents 5,445,942.

In another preferred embodiment, the protein of the invention or part thereof may be used to diagnose, treat and/or prevent disorders where the presence of substrates is undesirable or deleterious. Such disorders include but are not limited to infantile neuronal ceroid lipofuscinosis and lysosomal diseases. For diagnostic purposes, the expression of the protein of the invention
10 could be investigated using any of the Northern blotting, RT-PCR or immunoblotting methods described herein and compared to the expression in control individuals. For prevention and/or treatment purposes, the expression of protein of the invention may be enhanced using any of the gene therapy methods described herein or known to those skilled in the art.

In addition, the protein of the invention or part thereof may be used to identify inhibitors for
15 mechanistic and clinical applications. Such inhibitors may then be used to identify or quantify the protein of the invention in a sample, and to diagnose, treat or prevent any of the disorders where the protein's hydrolytic activity is undesirable and/or deleterious including but not limited to lysosomal diseases, neurodegenerative disorder such as infantile neuronal ceroid lipofuscinosis, Parkinson's and Alzheimer's diseases, inflammatory and immune disorders including allergies and leukemia.

20 Another object of the present invention are compositions and methods of targeting heterologous compounds, either polypeptides or polynucleotides to lysosomes by recombinantly or chemically fusing a fragment of the protein of the invention to an heterologous polypeptide or polynucleotide. Preferred fragments are any fragments of the protein of the invention, or part thereof, that may contain targeting signals for lysosomes such as those described in Vitale *et al.*,
25 *Mol.Cell.Biol.*, **20**: 7342-52 (2000), Blagoveshchenskaya *et al.*, *J.Biol.Chem.*, **273**: 2729-37 (1998) and Kornfeld, *FASEB J.*, **1**: 462-8 (1987)). Such heterologous compounds may be used to modulate lysosomal activity. For example, they may be used to induce and/or prevent a lysosomal protein degradation. Moreover, antibodies binding to the protein of the invention or part thereof may be used for detection of the lysosomes using any techniques known to those skilled in the art.

30 In still another embodiment, the invention relates to methods and compositions using the protein of the invention or part thereof as a marker protein to selectively identify tissues, preferably brain tissues. For example, the protein of the invention or part may be used to synthesize specific antibodies using any techniques known to those skilled in the art including those described therein. Such tissue-specific antibodies may then be used to identify tissues of unknown origin, for example,
35 forensic samples, differentiated tumor tissue that has metastasized to foreign bodily sites, or to differentiate different tissue types in a tissue cross-section using immunochemistry.

Another embodiment of the present invention relates to methods and compositions using the protein of the invention or part thereof to modify plant lipid composition using any assay known to those skilled in the art including those described by the US patents 5,955,650, 5,945,585 and 5,807,893. Indeed, plant lipids have a variety of nutritional uses and many recent research efforts have examined the role that saturated and unsaturated fatty acids play in reducing the risk of coronary heart disease. In the past, it was believed that mono-unsaturates, in contrast to saturates and poly-unsaturates, had no effect on serum cholesterol and coronary heart disease risk. Several recent human clinical studies suggest that diets high in mono-unsaturated fat and low in saturated fat may reduce the "bad" (low-density lipoprotein) cholesterol while maintaining the "good" (high-density lipoprotein) cholesterol (Mattson *et al.*, *Journal of Lipid Research*, 26: 194-202 (1985)).

In still another embodiment, the protein of the invention or part thereof may be used in enzyme replacement therapy, due to the ability of cells to take up exogenously supplied protein and target it to lysosomes (Neufeld E.F., *Annu.Rev.Biochem.* 60: 257-280(1991), Brady R.O. *et al*, *J.Inher.Metab.Dis.* 17: 510-519 (1994)), or in bone-marrow transplantation (Hoogerbrugge P.M. *et al*, *Lancet*, 345: 1398-1402 (1995)), as bone-marrow-derived microglial cells are believed to penetrate the blood-brain barrier and may theoretically be able to provide sufficient enzyme to correct the metabolic defect in neurons (Krivit W., *Cell transplant.*, 4: 385-392 (1995)). The protein of the invention or part thereof may be also used in genetic engineering of transplanted cells (Salvetti A. *et al*, *Br.Med.J.* 51: 106-122 (1995)) or neural progenitor cell engraftment (Snyder E.Y., *Nature*, 374: 367-370 (1995)) using any technique known to those skilled in the art.

Protein of SEQ ID NO: 254 (internal designation 106-006-1-0-E3-CS)

Angiogenin is a member of the pancreatic Rnase superfamily of proteins. Its mechanism of action is postulated to involve multiple interactions with other proteins through specific regions on the molecular surface of angiogenin. Potential partners of angiogenin include heparin, plasminogen, elastase, angiostatin, actin, and a 170 kDa receptor on the surface of endothelial cells [Strydom, D. J. (1998) *Cell. Mol. Life Sci.* 54, 811-824].

Angiogenin is required for the process of angiogenesis. Tumor growth requires angiogenesis, and several anti-angiogenic agents have been produced and are currently in the clinical trial stage. It has also been shown that recurrent gastric cancer patients had a much higher serum concentration of angiogenin than primary gastric cancer patients [Shimoyama, S. and Kaminishi, M. (2000) *J. Cancer Res. Clin. Oncol.* 126, 468-474]. Therefore, angiogenin can be used as a diagnostic marker for the evaluation of cancer aggressiveness or as an early marker for recurrence over a follow-up period.

Angiogenin is a potent inducer of angiogenesis [Fett, J. W.; Strydom, D. J.; Lobb, R. R.; Alderman, E. M.; Bethune, J. L.; Riordan, J. F.; and Vallee, B. L. (1985) *Biochemistry* 24, 5480-5486]. Angiogenesis is a complex process of blood vessel formation comprising of several separate

but interconnected steps at the cellular and biochemical level including: (i) activation of endothelial cells by the action of an angiogenic stimulus, (ii) adhesion and invasion of activated endothelial cells into the surrounding tissues and migration toward the source of the angiogenic stimulus, and (iii) proliferation and differentiation of endothelial cells to form a new microvasculature [Folkman, J. and Shing, Y. (1992) *J. Biol. Chem.* 267, 10931-10934; Moscatelli, D. and Rifkin, D. B. (1988) *Biochim. Biophys. Acta* 948, 67-85].

Angiogenin has been demonstrated to induce most of the individual events in the process of angiogenesis including binding to endothelial cells [Badet, J.; Soncin, F.; Guitton, J.D.; Lamare, O.; Cartwright, T.; and Barritault, D. (1989) *Proc. Natl. Acad. Sci. U.S.A.* 86, 8427-8431], stimulating second messengers [Bicknell, R. and Vallee, B. L. (1988) *Proc. Natl. Acad. Sci. U.S.A.* 85, 5961-5965], mediating cell adhesion [Soncin, F. (1992) *Proc. Natl. Acad. Sci. U.S.A.* 89, 2232-2236], activating cell-associated proteases [Hu, G. F. and Riordan, J. F. (1993) *Biochem. Biophys. Res. Commun.* 197, 682-687], inducing cell invasion [Hu, G-F.; Riordan, J. F.; and Vallee, B. L. (1994) *Proc. Natl. Acad. Sci. U.S.A.* 91, 12096-12100], inducing proliferation of endothelial cells [Hu, G-F.; Riordan, J. F.; and Vallee, B. L. (1997) *Proc. Natl. Acad. Sci. U.S.A.* 94, 2204-2209] and organizing the formation of tubular structures from the cultured endothelial cells [Jimi, S-I.; Ito, K-I.; Kohno, K.; Ono, M.; Kuwano, M.; Itagaki, Y.; and Isikawa, H. (1985) *Biochem. Biophys. Res. Commun.* 211, 476-483]. Angiogenin has also been shown to undergo nuclear translocation in endothelial cells via receptor-mediated endocytosis [Moroianu, J. and Riordan, J. F. (1994) *Proc. Natl. Acad. Sci. U.S.A.* 91, 1677-1681] and nuclear localization sequence-assisted nuclear import [Moroianu, J. and Riordan, J. F. (1994) *Biochem. Biophys. Res. Commun.* 203, 1765-1772].

While angiogenesis is a tightly-controlled process under usual physiological conditions, abnormal angiogenesis can have devastating consequences in pathological conditions such as arthritis, diabetic retinopathy and tumor growth. It is now well-established that the growth of virtually all solid tumors is angiogenesis dependent [Folkman, J. (1989) *J. Natl. Cancer Inst.* 82, 4-6]. Angiogenesis is also a prerequisite for the development of metastasis, since it provides the means whereby tumor cells disseminate from the original primary tumor and establish at distant sites [Mahadevan, V. and Hart, I. R. (1990) *Rev. Oncol.* 3, 97-103; Blood, C. H. and Zetter B. R. (1990) *Biochim. Biophys. Acta* 1032, 89-118]. Therefore, interference with the process of tumor-induced angiogenesis can be an effective therapy for both primary and metastatic cancers.

Although originally isolated from medium conditioned by human colon cancer cells (Fett et al. (1985), *supra*), and subsequently shown to be produced by several other histological types of human tumors [Rybak, S. M.; Fett, J. W.; Yao, Q-Z.; and Vallee, B. L. (1987) *Biochem. Biophys. Res. Commun.* 146, 1240-1248; Olson, K. A.; Fett, J. W.; French, T. C.; Key, M. E.; and Vallee, B. L. (1995) *Proc. Natl. Acad. Sci. U.S.A.* 92, 442-446], angiogenin also is a constituent of human plasma and normally circulates at a concentration of 250-360 ng/ml [Shimoyama, S.; Gansauge, F.;

Gansauge, S.; Negri, G.; Oohara, T.; and Beger, H. G. (1996) *Cancer Res.* 56, 2703-2706; Blaser, J.; Triebel, S.; Kopp, C.; and Tschesche, H. (1993) *Eur. J. Clin. Chem. Clin. Biochem.* 31, 513-516].

Several inhibitors of the functions of angiogenin have been developed. These include: (i) monoclonal antibodies (mAbs) [Fett, J. W.; Olson, K. A.; and Rybak, S. M. (1994) *Biochemistry* 33, 5421-5427], (ii) an angiogenin-binding protein [Hu, G-F.; Chang, S-I.; Riordan, J. F.; and Vallee, B. L. (1991) *Proc. Natl. Acad. Sci. U.S.A.* 88, 2227-2231; Hu, G-F.; Strydom, D. J.; Fett, J. W.; Riordan, J. F.; and Vallee, B. L. (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90, 1217-1221; Moroianu, J.; Fett, J. W.; Riordan, J. F.; and Vallee, B. L. (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90, 3815-3819], (iii) the placental ribonuclease inhibitor (PRI) [Shapiro, R. and Vallee, B. L. (1987) *Proc. Natl. Acad. Sci. U.S.A.* 84, 2238-2241], (iv) peptides synthesized based on the C-terminal sequence of angiogenin [Rybak, S. M.; Auld, D. S.; St. Clair, D. K.; Yao, Q-Z.; and Fett, J. W. (1989) *Biochem. Biophys. Res. Commun.* 162, 535-543], and (v) inhibitory site-directed mutagenesis of angiogenin [Shapiro, R. and Vallee, B. L. (1989) *Biochemistry* 28, 7401-7408].

The subject invention provides the protein/polypeptide of SEQ ID NO: 254. The invention also provides biologically active fragments of SEQ ID NO: 254. In one embodiment, the polypeptides of SEQ ID NO: 254 are interchanged with the corresponding polypeptides encoded by the human cDNA of clone 106-006-1-0-E3-CS. "Biologically active fragments" are defined as those peptide or polypeptide fragments having at least one of the biological functions of the full length protein (e.g., stimulation of angiogenesis). Compositions of the protein/polypeptide of SEQ ID NO: 254, or biologically active fragments thereof, are also provided by the subject invention. These compositions may be made according to methods well known in the art.

The invention also provides variants of the protein of SEQ ID NO: 254. These variants have at least about 80%, more preferably at least about 90%, and most preferably at least about 95% amino acid sequence identity to the amino acid sequence encoded by SEQ ID NO: 254. Variants according to the subject invention also have at least one functional or structural characteristic of the protein of SEQ ID NO: 254. The invention also provides biologically active fragments of the variant proteins. Compositions of variants, or biologically active fragments thereof, are also provided by the subject invention. These compositions may be made according to methods well known in the art. Unless otherwise indicated, the methods disclosed herein can be practiced utilizing the protein encoded by SEQ ID NO: 254, biologically active fragments of SEQ ID NO: 254, variants of SEQ ID NO: 254, and biologically active fragments of the variants.

Because of the redundancy of the genetic code, a variety of different DNA sequences can encode the amino acid sequence of SEQ ID NO: 254. In a preferred embodiment, SEQ ID NO: 254 is encoded by clone 106-006-1-0-E3-CS. It is well within the skill of a person trained in the art to create these alternative DNA sequences which encode proteins having the same, or essentially the same, amino acid sequence. These variant DNA sequences are, thus, within the scope of the subject invention. As used herein, reference to "essentially the same" sequence refers to sequences that

have amino acid substitutions, deletions, additions, or insertions that do not materially affect biological activity. Fragments retaining one or more characteristic biological activity of the protein encoded by clone 106-006-1-0-E3-CS are also included in this definition.

"Recombinant nucleotide variants" are alternate polynucleotides which encode a particular protein. They can be synthesized, for example, by making use of the "redundancy" in the genetic code. Various codon substitutions, such as the silent changes which produce specific restriction sites or codon usage-specific mutations, can be introduced to optimize cloning into a plasmid or viral vector or expression in a particular prokaryotic or eukaryotic host system, respectively.

In one aspect of the subject invention, SEQ ID NO: 254, and variants thereof, can be used to generate polyclonal or monoclonal antibodies. Both biologically active and immunogenic fragments of SEQ ID NO: 254, or variant proteins, can be used to produce antibodies. Polyclonal and/or monoclonal antibodies can be made according to methods well known to the skilled artisan. Antibodies produced in accordance with the subject invention can be used in a variety of detection assays known to those skilled in the art. The antibodies may be used to agonize or antagonize the biological activity of the protein of SEQ ID NO: 254.

SEQ ID NO: 254 can be used as a marker for individuals at risk for the development or recurrence of tumors. As indicated supra, angiogenin is found at certain levels in normal individuals, normally at concentrations of 250-360 ng/ml. Thus, quantitative immunoassays can be used for the detection of abnormal levels of SEQ ID NO: 254, thereby identifying those individuals at risk for the development of tumors. Alternatively, the subject invention provides antibodies specific for SEQ ID NO: 254, or fragments thereof, which are used in routine immunoassays to screen for the presence or absence of SEQ ID NO: 254, or fragments thereof.

Alternatively, the nucleic acids which encode SEQ ID NO: 254, or fragments thereof, may be used in hybridization assays to detect and/or quantitate the expression of SEQ ID NO: 254. Such hybridization assays are well known to the skilled artisan and can be practiced on a variety of samples, including, but not limited to, tumor cells, biopsied tissues, or normal tissue.

Molecules (see Strydom, D. J., (1998) Cell. Mol. Life Sci. 54, 811-824) that functionally inhibit the action of angiogenin can be used to treat patients with tumors. Because angiogenin is required for the vascularization of tumors, molecules which inhibit the biological activity of angiogenin can be used to reduce tumor vascularization and control tumor growth. Thus, another aspect of the invention provides molecules which inhibit, or reduce, the biological activity of SEQ ID NO: 254. One embodiment provides neutralizing antibodies to inhibit the biological activity of SEQ ID NO: 254. These neutralizing antibodies may be chimeric or humanized, according to methods well known in the art, to minimize the immunogenicity of the molecules when used in patients. Neutralizing antibodies may be used in conjunction with other known therapeutic modalities for the treatment of tumors.

Another embodiment of the invention utilizes the concept that expression of specific genes can be suppressed by oligonucleotides having a nucleotide sequence complementary to the mRNA transcript of the target gene. This suppression occurs by selectively impeding translation and has been termed an "antisense" methodology. In addition, "antigene" or "triplex" methodologies may also suppress expression of genes by using an oligonucleotide which is complementary to a selected site of double stranded DNA, thereby forming a triple-stranded complex to selectively inhibit transcription of the gene. Both "antisense" and "antigene" methodologies can be used to inhibit or reduce the expression of the gene of SEQ ID NO: 254, and thereby provide therapeutic benefit to the patient being treated. Methods of treating individuals using antigene and antisense methodologies are well known to those skilled in the art (see, for example, "Antisense Therapeutics" Agrawal, S. (ed), Humana Press, 1996; Crooke, S. T., and Bennett, C. F. (1996) Annu. Rev. Pharmacol. Toxicol. 36, 107-129; "Prospects for the Therapeutic Use of Antigene Oligonucleotides", Maher, L. J. (1996) Cancer Investigation 14(1), 66-82 each hereby incorporated by reference in its entirety).

As additional examples, U.S. Pat. No. 5,098,890 is directed to antisense oligonucleotides complementary to the c-myc oncogene and antisense oligonucleotide therapies for certain cancerous conditions. U.S. Pat. No. 5,135,917 provides antisense oligonucleotides that inhibit human interleukin-1 receptor expression. U.S. Pat. No. 5,087,617 provides methods for treating cancer patients with antisense oligonucleotides. U.S. Pat. No. 5,166,195 provides oligonucleotide inhibitors of HIV. U.S. Pat. No. 5,004,810 provides oligomers capable of hybridizing to herpes simplex virus Vmw65 mRNA and inhibiting replication. U.S. Pat. No. 5,194,428 provides antisense oligonucleotides having antiviral activity against influenza virus. U.S. Pat. No. 4,806,463 provides antisense oligonucleotides and methods using them to inhibit HTLV-III replication. U.S. Pat. No. 5,286,717 is directed to a mixed linkage oligonucleotide phosphorothioates complementary to an oncogene. U.S. Pat. No. 5,276,019 and U.S. Pat. No. 5,264,423 are directed to phosphorothioate oligonucleotide analogs used to prevent replication of foreign nucleic acids in cells. Each of these patents is hereby incorporated by reference in its entirety.

The subject invention also provides modified/derivatized nucleic acids encoding SEQ ID NO: 254. These include those modifications which increase the stability and/or affinity of these compounds for targets. Phosphorothioate analogs of oligodeoxynucleotides (ODNs), in which nonbridging phosphoryl oxygens in the backbone of DNA are substituted with sulfur ([S]ODNs) are substantially more stable than their native phosphodiester counterparts. Other derivatives, such as those alkylated on sugar oxygen groups, show enhanced target affinity. [S]ODNs possess good biological activity, pharmacology, pharmacokinetics and safety *in vivo* (Agrawal (1996), *supra*). Successful inhibition of specific gene function has been achieved by targeting various sites on specific mRNA sequences that include the AUG translational initiation codon, 5'-transcriptional

start site, 3'-termination codon and sequences in both the 5' and 3'-untranslated regions. These derivatized nucleic acids can be used in any of the aforementioned methodologies.

Protein of SEQ ID: 387 (internal designation 105-073-2-0-A7-CS)

The protein of SEQ ID NO : 387 encoded by the cDNA of SEQ ID NO: 146 is expressed in liver, ovary, prostate and overexpressed in salivary glands. The protein of SEQ ID NO : 387 belongs to the abhydrolase family, and is characterized by the alpha/beta hydrolase fold (Protein Eng 1992;5:197-211, which disclosure is hereby incorporated by reference in its entirety), that is common to a number of hydrolytic enzymes of widely differing phylogenetic origin and catalytic function.

10 The core of each enzyme is an alpha/beta-sheet (rather than a barrel), containing 8 strand connected by helices. The enzymes are believed to have diverged from a common ancestor, preserving the arrangement of the catalytic residues. All have a catalytic triad, the elements of which are borne on loops, which are the best conserved structural features of the fold.

Epoxide hydrolases are a family of enzymes which hydrolyze a variety of exogenous and endogenous epoxides to their corresponding diols. The epoxide hydrolase add water to epoxides, forming the corresponding diol. On the basis of sequence similarity, it has been proposed that the mammalian soluble epoxide hydrolase contain 2 evolutionarily distinct domains, the N-terminal domain is similar to bacterial haloacid dehalogenase, while the C-terminal domain is similar to soluble plant epoxyde hydrolase, microsomal epoxide hydrolase, and bacterial haloalkane dehalogenase (DNA Cell Biol. 14 :61-71 (1995), which disclosure is hereby incorporated by reference in its entirety. Human epoxide hydrolase catalyse the addition of water to epoxides to form the corresponding dihydrodiol. The enzymatic hydration is essentially irreversible and produces mainly metabolites of lower reactivity that can be conjugated and excreted. The reaction of epoxide hydrolase is therefore generally regarded as detoxifying. Commonly the function of epoxide hydrolase is finally followed by excretion of the diols. However, reactivation of certain diols by a second epoxidation may happen. Epoxide hydrolase inactivates also the epoxides existing in the metabolism of endogenous compounds. Lipophilic xenobiotics tend to accumulate into tissues, and they must be transformed to water soluble compounds to enable the excretion. In this transformation process reactive intermediates are produced. If biotransformation fails to detoxify these reactive intermediates, they may react covalently with critical targets like the genetic material, or start harmful reaction chains like lipid peroxidation. Therefore, epoxide hydrolases are thought to be responsible for carcinogenicity and mutagenicity phenomenon (Exp Pathol 1990;39(3-4):195-6.). In addition, the interaction between epoxide hydrolase activity and alcohol-metabolizing enzymes, suggests that epoxide hydrolase activity may be associated with the susceptibility to alcoholic liver disease and hepatocellular carcinoma (Toxicol. Lett. 10 ;115 (1) :17-22 (2000), which disclosure is hereby incorporated by reference in its entirety). Compounds containing the epoxide functionality

have become common environmental contaminants because of their wide use as pesticides, sterilants, and industrial precursors. Such compounds also occur as products, by-products, or intermediates in normal metabolism and as the result of spontaneous oxidation of membrane lipids (i.e. see, Brash, et al., Proc. Natl. Acad. Sci., 85:3382-3386 (1988), and Sevanian, A., et al.,
5 Molecular Basis of Environmental Toxicology (Bhatnager, R. S., ed.) pp. 213-228, Ann Arbor Science, Michigan (1980)). As three-membered cyclic ethers, epoxides are often very reactive and have been found to be cytotoxic, mutagenic and carcinogenic (i.e. see Sugiyama, S., et al., Life Sci. 40:225-231 (1987)). Cleavage of the ether bond in the presence of electrophiles often results in adduct formation. As a result, epoxides have been implicated as the proximate toxin or mutagen for
10 a large number of xenobiotics. Reactions of detoxification using epoxide hydrolases typically decrease the hydrophobicity of a compound, resulting in a more polar and thereby excretable substance.

It is believed that the protein of SEQ ID NO: 387 or part thereof is an hydrolase, preferably an epoxyde hydrolase. Preferred polypeptides of the invention are polypeptides comprising the
15 amino acids of SEQ ID NO: 387 from positions 2 to 132, 52 to 137, 29 to 120, 12 to 137, 19 to 136, 151 to 209, 141 to 209, 30 to 108, and 35 to 108. Other preferred polypeptides of the invention are fragments of SEQ ID NO: 387 having any of the biological activity described herein. The hydrolytic activity of the protein of the invention or part thereof may be assayed using any of the assays known to those skilled in the art including those described in Cancer res 40(7):2552-6
20 (1980); Exp Pathol 39(3-4):195-6 (1990), which disclosures are hereby incorporated by reference in their entireties.

The invention also relates to methods and compositions using the protein of the invention or part thereof to diagnose, prevent and/or treat several disorders linked to overexpression of the protein of the invention including alcoholic liver disease, hepatocellular carcinoma, ovarian and
25 prostate cancers.

In addition, the protein of the invention or part thereof may be used to identify inhibitors for mechanistic and clinical applications. Such inhibitors may then be used to identify or quantify the protein of the invention in a sample, and to diagnose, treat or prevent any of the disorders where the protein's hydrolytic activity is undesirable and/or deleterious such as disorders characterized by
30 tissue degradation including but not limited to amyloidosis, colitis, lysosomal diseases, arthritis, muscular dystrophy, inflammation, tumor invasion, glomerulonephritis, parasite-borne infections, Alzheimer's disease, periodontal disease, and cancer metastasis.

In another embodiment, the invention relates to methods and compositions using the protein of the invention or part thereof as a marker protein to selectively identify tissues, preferably ovarian,
35 liver or prostate, more preferably salivary glands. For example, the protein of the invention or part may be used to synthesize specific antibodies using any techniques known to those skilled in the art. Such tissue specific antibodies may then be used to identify tissues of unknown origin, for example,

forensic samples, differentiated tumor tissue that metastasized to foreign bodily, or to differentiate different tissue types in a tissue cross-section using immunochemistry.

Protein of SEQ ID No: 398 (internal designation: 160-31-3-0-E4-CS)

The protein of SEQ ID No: 398 encoded by the cDNA of SEQ ID No: 157, is
5 overexpressed in fetal brain and shows homology with diverse hydrolases. The protein of the invention also displays a motif characteristic of isochorismatase proteins from positions 17 to 147. In addition, the protein of the invention is an alternatively spliced form of an unnamed human protein.

It is believed that the protein of SEQ ID NO: 398 or part thereof is an hydrolase, preferably
10 acting on ether bonds, more preferably an ether hydrolase. Preferred polypeptides of the invention are polypeptides comprising the amino acids of SEQ ID NO: 398 from positions 17 to 147. Other preferred polypeptides of the invention are fragments of SEQ ID NO: 398 having any of the biological activity described herein. The hydrolytic activity of the protein of the invention or part thereof may be assayed using any of the assays known to those skilled in the art including those
15 described in US patents 5,445,942; 5,445,956, 6,017,746 and 5,871,616 and in Rusnak et al, 1990; Biochemistry 29 1425-1435.

In another embodiment, the invention relates to methods and compositions using the protein of the invention or part thereof as a marker protein to selectively identify tissues, preferably fetal brain. For example, the protein of the invention or part may be used to synthesize specific
20 antibodies using any techniques known to those skilled in the art including those described therein. Such tissue-specific antibodies may then be used to identify tissues of unknown origin, for example, forensic samples, differentiated tumor tissue that has metastasized to foreign bodily sites, or to differentiate different tissue types in a tissue cross-section using immunochemistry.

Proteins of SEQ ID NOs: 260 and 265 (internal designation 116-004-3-0-A6-CS and 116-091-1-0-D9-CS respectively)
25

The protein of SEQ ID NO: 260 encoded by the cDNA SEQ ID NO: 19 and over expressed in liver and testis is an isoform of the protein of SEQ ID NO: 265 encoded by the cDNA SEQ ID NO: 24 over expressed in liver. Both proteins show homology to murine EPCS26 (Hemberger M. et al., Dev. Biol. 222, 158-169 (2000)) with Genbank accession number AF250838. The proteins of
30 SEQ ID NO: 260 and 265 contain a signal peptide (cleavage site at position 18) that could allow the export of the protein to the extracellular domain, the export to a cellular membrane or to define a particular subcellular localization. The cDNA encoding EPCS26 has been shown to be differentially expressed during the process of trophoblast invasion.

Implantation and placentation are key processes in mammalian embryonic development.
35 They physically connect the embryo to its mother and are critical for sufficient nutrient and gas exchange. The extraembryonic cell lineage is the first to differentiate in the developing conceptus,

reflecting the importance of this cell for the establishment of fetal-maternal connections. During murine development, the outer layer of blastocyst, the mural trophoctoderm, begins to differentiate into primary trophoblast giant cells on day 5 of gestation (e5). These cells invade the uterine epithelium and penetrate deeply into the stroma. At the same time, the polar trophoctoderm cells continue to proliferate and form the ectoplacental cone. On e7, the outer cells of the ectoplacental cone begin to differentiate into secondary trophoblast giant cells. The invasion of uterine stroma by these cells is critical for successful placentation (Cross et al., Science 266, 1508-1518 (1994)).

Trophoblast invasion triggers secretion of proteinases that degrade extracellular matrix molecules. Mouse trophoblasts have been shown to synthesize and secrete serine proteases, matrix metalloproteinases and cysteine proteinases. Invasion of the trophoblast is a highly controlled process. The decidua restricts invasion by secreting proteinases inhibitors. Proteinases and proteinases inhibitors have antagonistic functions in implantation and placentation which may be mirrored by the reciprocity of their expression patterns (Alexander et al Development 122, 1723-1736 (1996)).

During tumor invasion and metastasis, the degradation of the basement membranes is often accomplished by the proteinases implicated in implantation and normal trophoblast invasion (Strickland and Richards Cell 71, 355-357 (1992), Wilson et al. Proc. Natl. Acad. Sci. USA 94, 1402-1407 (1997)). Uncontrolled trophoblast invasion, as in choriocarcinomas, results in one of the most metastatic tumors known (Strickland and Richards Cell 71, 355-357 (1992)).

A deficient function of the protein of the invention could result in an uncontrolled trophoblast invasion, and like in choriocarcinomas results in one of the most metastatic tumors known (Strickland and Richards Cell 71, 355-357 (1992)).

It is believed that the proteins of SEQ ID NO: 260 and 265 or part thereof play a role in proteolysis, preferably during embryogenesis, more preferably during trophoblast invasion. The proteins of the invention or part thereof may act as secreted proteinases that degrade extracellular matrix molecules or at the contrary as proteinase inhibitors. Preferred polypeptides of the invention are polypeptides comprising the amino acids of SEQ ID NO: 260 from positions 7 to 122 and the amino acids of SEQ ID NO: 265 from positions 7 to 81. Other preferred polypeptides of the invention are fragments of SEQ ID NO: 260 and 265 having any of the biological activities described herein. The proteolytic activity of the proteins of the invention or part thereof may be assayed using any of the assays known to those skilled in the art including those described in US patent 6,069,229 and 5,861,267. The protease inhibitor activity of the proteins of the invention or part thereof may be assayed using any of the assays known to those skilled in the art and using methods for determining inhibition constants well known to those skilled in the art (see Fersht, ENZYME STRUCTURE AND MECHANISM, 2nd ed., W.H. Freeman and Co., New York, (1985))

In addition, the proteins of the invention or part thereof may be used to diagnose, treat or prevent any of the disorders characterized by undesirable and/or deleterious hydrolytic activity such as disorders characterized by tissue degradation including but not limited to amyloidosis, colitis, lysosomal diseases, arthritis, muscular dystrophy, inflammation, tumor invasion,

- 5 glomerulonephritis, parasite-borne infections, Alzheimer's disease, periodontal disease, cancer metastasis, and choriocarcinoma. For diagnostic purposes, the expression of the proteins of the invention could be investigated using any of the Northern blotting, RT-PCR or immunoblotting methods described herein and compared to the expression in control individuals. Alternatively, inhibitors for the proteins' activity may be developed and use to inhibit and/or reduce its activity
- 10 using any methods known to those skilled in the art. Overexpression of the proteins of the invention or part thereof may be achieved using any of the gene therapy method described herein.

In another embodiment, the invention relates to methods and compositions using the protein of the inventions or part thereof as a marker protein to selectively identify tissues, preferably liver and testis for the protein of SEQ ID NO: 260, preferably liver for the protein of SEQ ID NO: 265.

- 15 For example, the proteins of the invention or part may be used to synthesize specific antibodies using any techniques known to those skilled in the art including those described therein. Such tissue-specific antibodies may then be used to identify tissues of unknown origin, for example, forensic samples, differentiated tumor tissue that has metastasized to foreign bodily sites, or to differentiate different tissue types in a tissue cross-section using immunochemistry.

20 *Protein of SEQ ID NO: 265 (internal designation 116-088-4-0-A9-CS)*

- The protein of SEQ ID NO: 265 encoded by the cDNA of SEQ ID NO: 24 is overexpressed in testis and liver. This protein of the invention is homologous to the GdX protein, also named UBL4 (Toniolo et al., Proc Natl Acad Sci USA 1988;85:851-5), found in both human (GENPEPT accession number L44140) and mice species (GENPEPT accession number J04761). In addition,
- 25 the 174-amino-acid-long protein of SEQ ID NO: 265, which is similar in size to ubiquitin-like proteins, displays a pfam consensus domain from position 1 to 82 that is the hallmarks of ubiquitin family proteins.

- Ubiquitin is a protein of 76 amino acid residues, found in all eukaryotic cells, and which is extremely well conserved from protozoan to vertebrates (Jentsch et al. Trends Cell Biol
- 30 2000;10:335-42). It plays a key role in a variety of cellular processes, such as ATP-dependent selective degradation of cellular proteins, maintenance of chromatin structure, regulation of gene expression, stress response, ribosome biogenesis, cell-cycle progression, signal transduction, transcription and antigen presentation (Wilkinson et al. Annu Rev Nutr 1995;15:161-89). The first ubiquitin is covalently ligated to target proteins through an isopeptide linkage between the C-
- 35 terminal glycine residue of ubiquitin and an internal ϵ -amino group of lysine residue of the substrate. To generated an efficient proteasomal targeting signal, additional ubiquitin are linked to

the first one by isopeptide bonds, and form branched poly-ubiquitin complexes (Thrower et al. EMBO J 2000;19: 94-102). Covalent binding of ubiquitin to proteins marks them for subsequent degradation by a multicomponent enzymatic complex known as the 26S proteasome (Hershko et al. Annu Rev Biochem 1992;61:761-807).

- 5 The genes coding ubiquitin-like proteins fall into two separate classes (Hershko et al. Annu Rev Biochem 1992;61:761-807). Proteins of the first class are frequently designed as ubiquitin-like modifiers, or UBLs. They produce polyubiquitin molecules consisting of exact head to tail repeats of ubiquitin, with a variable number of repeats. These linear polymer of ubiquitin are linked covalently through peptide bonds between the C-terminal glycine residue and N-terminal lysine
- 10 residue of contiguous ubiquitin molecules. Proteins of the second class are habitually named as ubiquitin-domain proteins, or UDPs. These proteins bear a single domain of the N-terminal domain that is related to ubiquitin, fused to a C-terminal ribosomal domain consisting of 52 or 76-80 amino-acid residues (Finley et al. Nature 1989;338:394-401). These proteins are not conjugated to other proteins and function as an heterogeneous group of proteins. To date, this family includes RAD23,
- 15 DSK2, PLIC-1, PLIC-2/Chap1, XDRP1, BAG-1, BAT3/Chap2, Scythe, Parkin, UIP28, UBP6, Elongin B, and GdX. In addition, the protein of invention of SEQ ID NO: 265 clearly belongs to the UDPs family, as it displays a single ubiquitin N-terminal consensus domain, which is the hallmark of this protein family subset.

- UDPs participate to regulation of proteolysis through multiple mechanisms such as
- 20 interaction with catalytically active 26S proteasome for RAD23 (Schauber et al. Nature 1998;391:715-8), hPLIC-1 and hPLIC-2 (Kleijnen et al. Mol Cell 2000;6:409-19), and BAG-1 (Luders et al. J Biol Chem 2000;275:4613-7), removing ubiquitin from conjugates for UBP6 (Wyndham et al. Protein Sci 1997;8:1268-75) and negative regulation of multi-ubiquitin chain assembly for RAD23 (Ortolan et al. Nature cell Biol 2000; 2:601-8). In addition, an increasing body
- 25 of evidence indicates that some UDPs participate to other cellular functions as protein folding (Luders et al. J Biol Chem 2000;275:4613-7), apoptosis (Kaye et al. FEBS Lett 2000;467:348-55), and nucleotide-excision repair (de Laat et al. Genes Dev 1999;13:768-785). UDPs family proteins have been shown directly associated with pathogenesis of several diseases including xeroderma pigmentosum for RAD23 (Masutani et al. EMBO J 1994;13:1831-43), and Parkinson's disease for
- 30 parkin (Kitada et al. Nature 1998;392:605-8). In addition, involvement of ubiquitin-like proteins or abnormal ubiquitinated accumulation of proteins has been found in multiple human disorders. Most of them, but not all, involve nervous central system as Alzheimer's disease (van Leeuwen et al. Science 1998;279:242-7), diffuse Lewy body disease (Iseki et al. J Neurol Sci 1997;146:53-7), Huntington disease (Scherzinger et al. Cell 1997;90:549-58), and amyotrophic lateral sclerosis
- 35 (Leigh et al. Brain 1991;114:775-88). In most disorders, ubiquitinated-proteins accumulate within cells and form aggregates termed inclusion bodies that have characteristic appearance on histological examination. In addition, abnormal accumulation of ubiquitinated proteins has been

found in Von-Hippel Lindau disease (Kamura et al. Proc Natl Acad Sci USA. 2000;97:10430-5), and in liver of alcoholic hepatitis patients (Ohta et al. Lab Invest. 1988;59:848-56). Components of hepatocytes are released within the circulation in alcoholic hepatitis (Sorbi et al. Am J Gastroenterol 1999;94:1018-22)

5 It is believed that the protein of SEQ ID NO: 265 or part thereof plays a role in the regulation of proteolysis, preferably as a ubiquitin-like protein, more preferably as a ubiquitin-domain protein. In addition, the protein of the invention may play a role in protein folding, apoptosis and nucleotide-excision repair. Preferred polypeptides of the invention are polypeptides comprising the amino acids of SEQ ID NO: 265 from positions 1 to 82. Other preferred
10 polypeptides of the invention are fragments of SEQ ID NO: 265 having any of the biological activity described herein.

In an embodiment, the invention relates to compositions and methods using the protein of the invention or part thereof to remove, identify or inhibit contaminating proteases in a sample. Compositions comprising the polypeptides of the present invention may be added to biological
15 samples as a "cocktail" with other protease inhibitors to prevent degradation of protein samples. The advantage of using a cocktail of protease inhibitors is that one is able to inhibit a wide range of proteases without knowing the specificity of any of the proteases. Using a cocktail of protease inhibitors also protects a protein sample from a wide range of future unknown proteases which may contaminate a protein sample from a vast number of sources. Such protease inhibitor cocktails (see
20 for example the ready to use cocktails sold by Sigma) are widely used in research laboratory assays to inhibit proteases susceptible of degrading a protein of interest for which the assay is to be performed. For example, the protein of the invention or part thereof is added to samples where proteolytic degradation by contaminating proteases is undesirable. Alternatively, the protein of the invention or part thereof may be bound to a chromatographic support, either alone or in
25 combination with other protease inhibitors, using techniques well known in the art, to form an affinity chromatography column. A sample containing the undesirable protease is run through the column to remove the protease. Alternatively, the same methods may be used to identify new proteases.

Another embodiment of the invention relates to compositions and methods of using the
30 protein of invention or part thereof to develop assays for the immunohistochemical detection of testicular malignant tissue, as the protein is overexpressed in such tissue. For instance, this could be used for staging lymph node testicular cancer dissemination using the techniques and methods detailed in Nazeer et al. Oncol Rep (1998);5:1425-9. The ability to specifically visualize malignant tissues (and cells derived from the tissues), is useful for numerous applications, including to
35 determine the origin, to identify e.g. cancerous cells, as well as to facilitate the identification of particular cells and tissues for, e.g. the evaluation of histological slides.

In another embodiment, the invention relates to compositions or methods using the protein of SEQ ID NO: 265 or part thereof to diagnose, treat and/or prevent disorders including, but not limited to xeroderma pigmentosum, Von-Hippel Lindau disease, alcoholic hepatitis, in neurodegenerative diseases such as Alzheimer's disease, diffuse Lewy body disease, Huntington disease, and amyotrophic lateral sclerosis. Detection of poly-ubiquitinated protein conjugates in biological samples, such as brain tissues for the diagnosis of neurodegenerative disorders or liver and serum or plasma for the diagnosis of alcoholic hepatitis, may be performed using antibodies or nucleic acid able to detect the expression of the protein of the invention using immunohistochemistry, enzyme-linked immunosorbant assay (ELISA) or any other technique known to those skilled in the art including Northern blotting, RT-PCR or immunoblotting methods described herein as well as the technique described in Mimnaugh et al. Electrophoresis 1999;20:418-28. The expression of the protein of the invention in patients' samples is then compared to the expression in control individuals.

In still another embodiment, the invention relates to compositions or methods to treat, attenuate and/or prevent disorders including, but not limited to xeroderma pigmentosum, Von-Hippel Lindau disease, alcoholic hepatitis, in neurodegenerative diseases such as Alzheimer's disease, diffuse Lewy body disease, Huntington disease, and amyotrophic lateral sclerosis using the protein of the invention, part thereof, or any other compounds developed using the present protein as nucleic acids, antibodies, or chemical substances. In a preferred embodiment, proteins or other compounds targeted against the protein of invention or part thereof may be used to treat, prevent and/or attenuate disorders in which ubiquitin-like proteins or abnormal accumulation of ubiquitinated proteins has been found and can be involved in pathogenesis of the disease. For instance, proteins or other compounds targeted against protein of SEQ ID NO: 265 can be administered to treat or attenuate symptoms of patients affected with Alzheimer's disorder or any other neurodegenerative disorders.

Protein of SEQ ID NO: 408 (internal designation 174-8-2-0-C10-CS)

The protein of SEQ ID NO: 408 encoded by the cDNA of SEQ ID NO: 167 found in salivary gland and brain is homologous to a drosophila melanogaster protein thought to be transmembraneous (STR: Q9V641). The 345-amino-acid-long protein of SEQ ID NO: 408 displays the Rhomboid pfam domain from positions 186 to 323 and is predicted having six transmembrane domains from positions 101 to 121, 167 to 187, 204 to 224, 243 to 263, 273 to 293, 298 to 318.

Rhomboid genes were identified in flies and in organisms as diverse as Arabidopsis, yeast, bacteria, and mammals. Human and rat homologues of Rhomboid have been identified (Pascal et al.: 1998; FEBBS Lett. 429; 337-340). This very widespread conservation implies that the Rhomboid family proteins have a fundamental function within many cells. The Drosophila

Rhomboid has six transmembrane domains and an amino terminal hydrophobic region like the protein of the invention.

The 355-amino-acid-long *Drosophila* Rhomboid protein is known to control many aspects of fly development and especially, to establish position along the dorsoventral axis and then again later to specify the fate of neuronal precursor cells. Rhomboid expression is sufficient to activate EGF receptor (EGFr) signaling in all tissues in *Drosophila*, while loss of Rhomboid mimics reduction (or loss) of EGFr signaling in almost all tissues (Guichard et al.: 1999 Development; 126, 2663-2676). As in mammals, the *drosophila* EGF receptor controls many aspects of growth and development. Three activating ligands of the *drosophila* EGFr have been described, the most developmentally significant being the TGF alpha-like molecule, Spitz (Rutledge et al.: 1992; Genes & Dev. 6; 1503-1517). None of the Rhomboid-like proteins from species other than *Drosophila* have clearly assigned functions. However, there is compelling genetic evidence from *Drosophila* that Rhomboid has a key role in intercellular signaling: it functions as an activator of the EGF receptor, probably by controlling the activation the TGF-like ligand Spitz (Guichard et al.: 1999 Development; 126, 2663-2676). Indeed, Rhomboid expression is the principal rate-limiting step in activation of the Ras/MAP kinase pathway by the EGFr.

Like mammalian TGF alpha, Spitz is synthesized as a functionally inert transmembrane protein; subsequently, the proteolytic release of the extracellular portion of the molecule gives rise to a soluble and potent EGFr ligand (Golembo et al.: 1996; Development; 122; 3363-3370). Unlike all other essential components of EGFr signaling, the expression of Rhomboid is tightly restricted to sites of signaling activity. It has been proposed that Rhomboid attains its key role in the pathway by regulating the proteolytic cleavage of Spitz (Wasserman et al.: 2000; Genes & development; 14; 1651-1663). The preeminence of Rhomboid in a pathway as critical to development and growth control as the EGFr/Ras/Map kinase cascade provides a strong incentive to understand its molecular mechanism. By analogy to mammalian EGFr ligands that are similarly processed, Spitz cleavage is expected to be catalyzed by an ADAM like protease (Black et al 1998; Curr. Opin. Cell.Biol. 10; 654-659), but Rhomboid resembles no known protease.

The *Drosophila* eye has served as a useful model for studying mechanisms of EGFr and Ras signaling. At least five different roles for the receptor have been identified (for reviews see Wasserman et al.: 2000; Genes & development; 14; 1651-1663), the best characterized being its function in recruiting cells into the developing ommatidium- the individual unit of the fly compound eye. Each ommatidium contains eight photoreceptors, four cone cells that secrete lens material, and an average of eight pigment cells. It has been shown that the fly EGFr has a role in regulating cell survival in the developing eye (Dominguez et al. 1998; Curr. Biol. 8; 1039-1048).

The EGFr signaling pathway has been conserved between flies and vertebrates. The EGFr family consists of four members, HER1 (c-erbB1, EGFR), HER2 (c-erbB2), HER3 (c-erbB3), HER4 (c-erbB4), expressed in a wide range of cells (Gullick W.J. 1998; Br. Cancer Res. Treat.; 52,

43-53). TGF.alpha. and its homologs have been found to be the most abundant ligands for the EGF/TGF.alpha. receptor in most parts of the brain (Kaser, et al., (1992)Brain Res Mol Brain Res: 16:316-322). There appears to be a widespread distribution of TGF.alpha. in various regions of the brain in contrast to EGF which is only present in smaller, more discrete areas, suggesting that TGF-alpha might play a physiological role in brain tissues. These numerous receptor sites for TGF.alpha in the brain suggest that TGF has an important utility in promoting normal brain cell differentiation and function.

Transforming growth factor alpha (TGF-alpha.) is a relative of epidermal growth factor (EGF) and like EGF, it exerts its effects on cells through binding to the EGF receptor. The precise physiological roll of TGF.alpha. is still not clear, although it appears to be important in eye and hair follicle development and may play a role in both the immune system and in wound healing. (See Kumar, et al.; 1995 Cell Biology International, 19:5, 373-388). The EGF family receptors currently includes four EGF receptors. The EGFR2 receptor may also be referred to as ERB-2 and this molecule is useful for a variety of diagnostic and therapeutic indications (Prigent, S. A., and Lemoine, N. R., (1992) Prog Growth Factor Res., 4:1-24). The TGF-alpha is likely a ligand for one or more of these receptors as well as for yet an identified new EGF-type receptor. Use of the TGF-alpha. can assist with the identification, characterization and cloning of such receptors. For example, the EGF receptor gene represents the cellular homolog of the v-erb-B oncogene of avian erythroblastosis virus. Over expression of the EGF-receptor or deletion of kinase regulatory segments of the protein can bring about tumorigenic transformation of cells (Manjusri, D. et al., (1991) Human Cytokines, 364 and 381).

The EGF receptor, and the related ErbB family of receptor tyrosine kinases, have indeed been much implicated in human cancer. It is commonly believed that hyperactive receptor signaling promotes dysregulates growth control and in involved in the onset of malignancy, as well as in the disruption of developmental programs. Very little, however, is known about ErbB physiological regulation in humans. The fruitfly, *Drosophila melanogaster*, has a single receptor homologous to the four ErbB receptors. As signaling mechanisms have been well conserved between flies and mammals, these results of experiments in flies are relevant to the study of the human receptors in development and disease. Two areas of recent progress are emphasized. First, a number of signal modulators have been identified, including three EGF receptor inhibitors, several of which have human homologues. Second, the signaling molecules are integrated into regulatory networks that specify the elaborate activation profiles needed in development (positive and negative feedback control of EGF receptor signaling emerges as a central theme).

It is thus important to discover whether Rhomboid-like proteins also have functions similar to those observed in *Drosophila* in other higher organisms, including mammals, because of the substantial clinical importance of the EGFr pathway.

It is believed that the protein of SEQ ID NO: 408 or part thereof plays a role into the control of cellular signaling. Preferably the protein of the invention or part thereof plays a role in the activation of EGFr-mediated cell signaling, probably through the control of the activation of EGFr ligands, such as EGF, TGF alpha and TGF alpha-like factor, more probably through the proteolytic cleavage of such ligands. Preferred polypeptides of the invention are polypeptides comprising the amino acids of SEQ ID NO: 408 from positions 186 to 323, 101 to 121, 167 to 187, 204 to 224, 243 to 263, 273 to 293, and 298 to 318. Other preferred polypeptides of the invention are fragments of SEQ ID NO: 408 having any of the biological activity described herein. The proteolytic activity of the protein of the invention or part thereof as well as its involvement in regulation of cellular signalling though the activation of EGFr may be assayed using any of the assays known to those skilled in the art.

An embodiment of the invention relates to composition and methods using the protein of the invention or part thereof to identify and/or quantify the activation of EGF receptors, preferably vertebrate EGF receptors, more preferably human ErbB receptors, in a biological sample, and thus used in assays and diagnostic kits for the quantification of such activation in bodily fluids, in tissue samples, and in mammalian cell cultures. The assessment of the activation of EGF receptors may be performed using any assay familiar to those skilled in the art. Preferably, a defined quantity of the protein of the invention or part thereof is added to the sample under conditions allowing the activation of EGFr. Then, the activation of EGFr is assayed and eventually compared to a control using any of the techniques known by those skilled in the art.

The present invention also relates to diagnostic assays for detecting altered levels of the protein of the present invention in various tissues since an over-expression of the proteins compared to normal control tissue samples can detect the presence of certain disease conditions such as neoplasia, skin disorders, ocular disorders and inflammation. Assays used to detect levels of the polypeptide of the present invention in a sample derived from a host are well-known to those of skill in the art and include radioimmunoassays competitive-binding assays, Western Blot analysis and preferably ELISA assays.

This invention is also related to the use of SEQ ID No: 167 or its complement as a diagnostic tool. Detection of a mutated form of the nucleotide sequence of SEQ ID No: 167 of the present invention will allow a diagnosis of a disease or a susceptibility to a disease which results from underexpression of the polypeptide of the present invention for example, improper wound healing, improper neurological functioning, ocular disorders, kidney and liver disorders, hair follicular development, angiogenesis and embryogenesis. Individuals carrying mutations in the human nucleotide sequence of SEQ ID No: 167 of the present invention may be detected at the DNA level by a variety of techniques. Nucleic acids for diagnosis may be obtained from a patient's cells, such as from blood, urine, saliva, tissue biopsy and autopsy material. The genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR (Saiki et al., (1986)

Nature, 324:163-166) prior to analysis. RNA or cDNA may also be used for the same purpose. As an example, PCR primers complementary to the nucleic acid encoding a polypeptide of the present invention can be used to identify and analyze mutations thereof. For example, deletions and insertions can be detected by a change in size of the amplified product in comparison to the normal
5 genotype. Point mutations can be identified by hybridizing amplified DNA to radiolabeled RNA or alternatively, radiolabeled antisense DNA sequences. Perfectly matched sequences can be distinguished from mismatched duplexes by RNase A digestion or by differences in melting temperatures.

In another embodiment, the protein of the invention or part thereof can be used to diagnose,
10 treat and/or prevent disorders linked to dysregulation of growth control, such as cancer and other disorders relating to abnormal cellular differentiation, proliferation, or degeneration, including hyperaldosteronism, hypocortisolism (Addison's disease), hyperthyroidism (Grave's disease), hypothyroidism, colorectal polyps, gastritis, gastric and duodenal ulcers, ulcerative colitis, and Crohn's disease, neurodegenerative disorders such as Parkinson's and Alzheimer's diseases using
15 any methods and/or techniques described herein. For diagnostic purposes, the expression of the protein of the invention could be investigated using any of the Northern blotting, RT-PCR or immunoblotting methods described herein and compared to the expression in control individuals. In addition, the protein of the invention or part thereof may be used to evaluate the disease progression and the clinical treatment efficiency. Inhibition of expression of the protein of the invention or part
20 thereof to inhibit EGFr activation could be achieved by many means known to those skilled in the art including those described in the present application such as antisense nucleotide or triple helix strategies.

Protein of SEQ ID NO:291 (internal designation: 180-19-4-0-F4-CS)

The protein of SEQ ID No:291 encoded by the cDNA of SEQ ID No:50 is homologous to
25 proteins of the tissue inhibitor of metalloproteinases (TIMP) family. The protein of the invention (207 amino-acids) is highly homologous to and appears to be a variant of the metalloproteinase inhibitor 1 precursor (TIMP-1, 207 amino-acids) human protein (SwissProt P01033). The protein of the invention is strongly expressed in the liver, ovary and testis.

There are many different types of collagen found in the body and they, together with other
30 extracellular matrix components, such as elastin, gelatin, proteoglycan and fibronectin, make up a large proportion of the body's extracellular tissue. Matrix metalloproteinases (MMPs) are enzymes that are involved in the degradation and denaturation of extracellular matrix components. Collagenases, for example, are MMPs that degrade or denature collagen. A large number of different collagenases are known to exist. These include interstitial collagenases, type IV-specific
35 collagenases and collagenolytic proteinases. Collagenases are generally specific for collagens which, in their full triple helix structure, are extremely resistant to other enzymes. Other MMPs are

involved in the degradation and denaturation of different extracellular matrix components, for example, elastin, gelatin and proteoglycan. Some MMPs are able to degrade or denature several different types of collagen and also other extracellular matrix components. For example, stromelysin degrades type IV collagen, which is found in basement membrane, and also has an effect on other extracellular matrix components such as elastin, fibronectin and cartilage proteoglycans. The ability of MMPs metalloproteinases (such as collagenase, stromelysin, and gelatinase) to degrade various components of connective tissue makes them potential targets for controlling numerous pathological processes.

The presence of tissue inhibitors of MMPs has been observed in a variety of explants and in monolayer cultures of mammalian connective tissue cells (Vater et al 1979 and Stricklin and Wegus 1983). Not only collagenase inhibitors but also inhibitors for other MMPs, for example, gelatinase and proteoglycanase have been found. MMP inhibitors are generally unable to bind the inactive (zymogen) forms of the respective enzymes but complex readily with active forms (Murphy et al 1981). Tissue MMP inhibitors are found, for example, in dermal fibroblasts, human lung, gingival, tendon and corneal fibroblasts, human osteoblasts, uterine smooth muscle cells, alveolar macrophages, amniotic fluid, plasma, serum and the α -granule of human platelets (Stricklin and Wegus 1983; Welgus et al 1985; Welgus and Stricklin 1983; Bar-Sharvit et al 1985; Wooley et al; 1976; and Cooper et al 1985).

The protein of the invention is a secreted TIMP-1 protein which tightly complexes with metalloproteinases and irreversibly inactivate them. TIMP-1 has been identified as a secretory product of platelets and alveolar macrophages

Thus, an embodiment of the present invention relates to the use of the protein of the invention or a fragment thereof to inhibit the action of MMPs by directly inhibiting the enzyme activity like a conventional inhibitor. The inhibitory activity of a MMP inhibitor may be assessed by any method suitable for determining inhibitory activity of a compound with respect to an enzyme. Such methods are described in standard textbooks of biochemistry.

In one embodiment, the protein of SEQ ID NO:291 can be used to treat and diagnose disorders associated with excessive MMP expression, such as inflammatory disorders such as rheumatoid arthritis, osteoarthritis, osteopenias such as osteoporosis, pulmonary emphysema, periodontitis, gingivitis, corneal epidermal or gastric ulceration, and tumour metastasis, invasion and growth, Paget's disease, hyperparathyroidism. MMP inhibitors are also of potential value in the treatment of neuroinflammatory disorders, including those involving myelin degradation, for example multiple sclerosis, as well as in the management of angiogenesis dependent diseases, which include arthritic conditions and solid tumour growth as well as psoriasis, proliferative retinopathies, neovascular glaucoma, ocular tumours, angiofibromas and hemangiomas. The present invention relates to a method of treating diseases in which MMPs are involved such as atherosclerotic plaque rupture, restenosis, aortic aneurysm (including abdominal aortic aneurysm

and brain aortic aneurysm), congestive heart failure, left ventricular dilatation, myocardial infarction, decubital ulcers, chronic ulcers or wounds, renal disease, or other autoimmune or inflammatory diseases dependent upon tissue invasion by leukocytes, Crohn's disease, acute respiratory distress syndrome, asthma, chronic obstructive pulmonary disease, Alzheimer's disease, organ transplant toxicity, cachexia, allergic reactions, allergic contact hypersensitivity, epidermolysis bullosa, loosening of artificial joint implants, stroke, cerebral ischemia, head trauma, spinal cord injury, neuro-degenerative disorders (acute and chronic), Huntington's disease, Parkinson's disease, migraine, depression, peripheral neuropathy, pain, cerebral amyloid angiopathy, nootropic or cognition enhancement, amyotrophic lateral sclerosis, ocular angiogenesis, macular degeneration, abnormal wound healing, burns, diabetes, scleritis, AIDS, sepsis, septic shock.

In another embodiment, the protein of SEQ ID NO:291 has potential value in the treatment or diagnosis of atherosclerosis. The rupture of atherosclerotic plaques is the most common event initiating coronary thrombosis. Destabilization and degradation of the extracellular matrix surrounding these plaques by MMPs has been proposed as a cause of plaque fissuring. The shoulders and regions of foam cell accumulation in human atherosclerotic plaques show locally increased expression of gelatinase B, stromelysin-1, and interstitial collagenase. In situ zymography of this tissue revealed increased gelatinolytic and caseinolytic activity (Galla, et al., J. Clin. Invest., 1994;94:2494-2503). In addition, high levels of stromelysin RNA message have been found to be localized to individual cells in atherosclerotic plaques removed from heart transplant patients at the time of surgery (Henney, et al., Proc. Nat'l. Acad. Sci., 1991;88:8154-8158).

In another embodiment, the protein of the invention has utility in treating or detecting degenerative aortic disease associated with thinning of the medial aortic wall. Increased levels of the proteolytic activities of MMPs have been identified in patients with aortic aneurysms and aortic stenosis (Vine N. and Powell J. T., Clin. Sci., 1991;81:233-239).

In another embodiment, the protein of the invention can be used as a treatment or diagnostic tool for heart failure and associated ventricular dilatation. Heart failure arises from a variety of diverse etiologies, but a common characteristic is cardiac dilation which has been identified as an independent risk factor for mortality (Lee, et al., Am. J. Cardiol., 1993;72:672-676). This remodeling of the failing heart appears to involve the breakdown of extracellular matrix. MMPs are increased in patients with both idiopathic and ischemic heart failure (Reddy, et al., Clin. Res., 1993;41:660A; Tyagi S. C., et al., Clin. Res., 1993;41:681A). Animal models of heart failure have shown that the induction of gelatinase is important in cardiac dilation (Armstrong, et al., Can. J. Cardiol., 1994;10:214-220), and cardiac dilation precedes profound deficits in cardiac function (Sabbah, et al., Am. J. Physiol., 1992;263:H266-H270).

In another embodiment, the protein of the invention is useful in treating or detecting neointimal proliferation, leading to restenosis, frequently developed after coronary angioplasty.

The migration of vascular smooth muscle cells (VSMCs) from the tunica media to the neointima is a key event in the development and progression of many vascular diseases and a highly predictable consequence of mechanical injury to the blood vessel (Bendeck M. P., et al., *Circulation Research*, 1994;75:539-545). Northern blotting and zymographic analyses indicated that gelatinase A was the principal MMP expressed and excreted by these cells. Further, antisera capable of selectively neutralizing gelatinase A activity also inhibited VSMC migration across basement membrane barrier. (Pauly R. R., et al., *Circulation Research*, 1994;75:41-54).

In another embodiment, the protein of the invention is used to ensure normal kidney function, which is dependent on the maintenance of tissues constructed from differentiated and highly specialized renal cells. Those cells are in a dynamic balance with their surrounding extracellular matrix (ECM) components (Davies M. et al., *Kidney Int.*, 1992;41:671-678). Effective glomerular filtration requires that a semi-permeable glomerular basement membrane (GBM) composed of collagens, fibronectin, enactin, laminin and proteoglycans is maintained. A structural equilibrium is achieved by balancing the continued deposition of ECM proteins with their degradation by specific MMPs. These proteins are first secreted as proenzymes and are subsequently activated in the extracellular space. These proteinases are in turn subject to counter balancing regulation of their activity by naturally occurring inhibitors as TIMPs.

Deficiency or defects in any component of the filtration barrier may have catastrophic consequences for longer term renal function. For example, in hereditary nephritis of Alport's type, associated with mutations in genes encoding ECM proteins, defects in collagen assembly lead to progressive renal failure associated with splitting of the GBM and eventual glomerular and interstitial fibrosis. In contrast, in inflammatory renal diseases such as glomerulonephritis, cellular proliferation of components of the glomerulus often precede obvious ultrastructural alteration of the ECM matrix. Cytokines and growth factors implicated in proliferative glomerulonephritis such as interleukin-1, tumor necrosis factor, and transforming growth factor beta can upregulate metalloproteinase expression in renal mesangial cells (Martin J. et al., *J. Immunol.*, 1986;137:525-529; Marti H. P. et al., *Biochem. J.*, 1993;291:441-446; Marti H. P. et al., *Am. J. Pathol.*, 1994;144:82-94). These metalloproteinases are believed to be intimately involved in the aberrant tissue remodeling and cell proliferation characteristic of renal diseases, such as, IgA nephropathy which can progress to through a process of gradual glomerular fibrosis and loss of functional GBM to end-stage renal disease. Metalloproteinase expression has already been well-characterized in experimental immune complex-mediated glomerulonephritis such as the anti-Thy 1.1 rat model (Bagchus W. M., et al., *Lab. Invest.*, 1986;55:680-687; Lovett D. H., et al., *Am. J. Pathol.*, 1992;141:85-98).

In another embodiment, the protein of the invention can be used as a treatment or diagnostic tool for gingiva. Collagenase and stromelysin activities have been demonstrated in fibroblasts isolated from inflamed gingiva (Uitto V. J., et al., *J. Periodontal Res.*, 1981;16:417-424), and

enzyme levels have been correlated to the severity of gum disease (Overall C. M., et al., J. Periodontal Res., 1987;22:81-88).

In another embodiment, the protein of the invention is useful for treating or detecting ulcers. Proteolytic degradation of extracellular matrix has been observed in corneal ulceration following alkali burns (Brown S. I., et al., Arch. Ophthalmol., 1969;81:370-373). Thiol-containing peptides inhibit the collagenase isolated from alkali-burned rabbit corneas (Burns F. R., et al., Invest. Ophthalmol., 1989;30:1569-1575). Stromelysin, a member of the MMP family, is produced by basal keratinocytes in a variety of chronic ulcers (Saarialho-Kere U. K., et al., J. Clin. Invest., 1994;94:79-88). Stromelysin-1 mRNA and protein were detected in basal keratinocytes adjacent to but distal from the wound edge in what probably represents the sites of proliferating epidermis. Stromelysin-1 may thus prevent the epidermis from healing.

In another embodiment, the protein of the invention can be used as a treatment or diagnostic tool for tumor angiogenesis. Inhibitors of MMPs have shown activity in models of tumor angiogenesis (Taraboletti G., et al., Journal of the National Cancer Institute, 1995;87:293; and Benelli R., et al., Oncology Research, 1994;6:251-257). Davies et al., (Cancer Res., 1993;53:2087-2091) reported that a peptide decreased the tumor burden and prolonged the survival of mice bearing human ovarian carcinoma xenografts. A peptide of the conserved MMP propeptide sequence was a weak inhibitor of gelatinase A and inhibited human tumor cell invasion through a layer of reconstituted basement membrane (Melchiori A., et al., Cancer Res., 1992;52:2353-2356), and the natural tissue inhibitor of metalloproteinase-2 (TIMP-2) also showed blockage of tumor cell invasion in in vitro models (DeClerck Y. A., et al., Cancer Res., 1992;52:701-708). Studies of human cancers have shown that gelatinase A is activated on the invasive tumor cell surface (Strongin A. Y., et al., J. Biol. Chem., 1993;268:14033-14039) and is retained there through interaction with a receptor-like molecule (Monsky W. L., et al., Cancer Res., 1993;53:3159-3164).

In another embodiment, the protein of the invention can be used to treat and diagnose rheumatoid arthritis. Collagenases have been implicated in a number of diseases, including, rheumatoid arthritis (Mullins, D. E. et al 1983), and it has been proposed to use MMP inhibitors in the treatment of this condition. Several investigators have demonstrated consistent elevation of stromelysin and collagenase in synovial fluids from rheumatoid and osteoarthritis patients as compared to controls (Walakovits L. A., et al., Arthritis Rheum., 1992;35:35-42; Zafarullah M., et al., J. Rheumatol., 1993;20:693-697). TIMP-1 and TIMP-2 prevented the formation of collagen fragments, but not proteoglycan fragments, from the degradation of both the bovine nasal and pig articular cartilage models for arthritis, while a synthetic peptide hydroxamate could prevent the formation of both fragments (Andrews H. J., et al., Biochem. Biophys. Res. Commun., 1994;201:94-101).

In another embodiment, the protein of the invention is used to treat or diagnose inflammation. Gijbels et al., (J. Clin. Invest., 1994;94:2177-2182) recently described a peptide that

suppressed the development or reversed the clinical expression of experimental allergic encephalomyelitis (EAE) in a dose dependent manner, suggesting the use of MMP inhibitors in the treatment of autoimmune inflammatory disorders such as multiple sclerosis. A recent study by Madri has elucidated the role of gelatinase A in the extravasation of T-cells from the blood stream
5 during inflammation (Ramanic A. M. and Madri J. A., J. Cell Biology, 1994;125:1165-1178). This transmigration past the endothelial cell layer is coordinated with the induction of gelatinase A and is mediated by binding to the vascular cell adhesion molecule-1 (VCAM-1). Once the barrier is compromised, edema and inflammation are produced in the CNS. Leukocytic migration across the blood-brain barrier is known to be associated with the inflammatory response in EAE. Inhibition of
10 the metalloproteinase gelatinase A would block the degradation of extracellular matrix by activated T-cells that is necessary for CNS penetration. These studies provided the basis for the belief that an inhibitor of stromelysin-1 and/or gelatinase A will treat diseases involving disruption of extracellular matrix resulting in inflammation due to lymphocytic infiltration, inappropriate migration of metastatic or activated cells, or loss of structural integrity necessary for organ function.

15 The present invention provides the use of an MMP inhibitor in the manufacture of a medicament for the treatment or prophylaxis of scars. Collagen is the major component of scar and other contracted tissue and as such is the most important structural component to consider. Contraction of tissues comprising extracellular matrix components, especially of collagen-comprising tissues, may occur in connection with many different pathological conditions and with
20 surgical or cosmetic procedures. Contracture, for example, of scars, may cause physical problems, which may lead to the need for medical treatment, or it may cause problems of a purely cosmetic nature.

During experiments on in vitro models of scar contraction, collagen appears to be invaded and permanently remodelled by fibroblasts and that such invasion and remodelling is inhibited by
25 collagenase inhibitors. The remodelling generally appears as contraction of the collagen, the contraction of which is inhibited by inhibition of collagenase. Furthermore, inhibition of other MMPs also results in inhibition of contraction.

The present invention also provides the use of an MMP inhibitor in the treatment or prophylaxis of a natural or artificial tissue comprising extracellular matrix components to inhibit,
30 i.e. restrict, hinder or prevent, contraction of the tissue, especially contraction resulting from a pathological condition or from surgical or cosmetic treatment.

Cosmetic treatments, such as chemical or physical dermal abrasion, used as anti-ageing treatments, cause trauma to the skin. Use of MMP inhibitors during the healing process which occurs after the initial abrasion is a cosmetic use of MMP inhibitors according to the present
35 invention.

The present invention also provides the use of an MMP inhibitor to inhibit, i.e. restrict, hinder or prevent, invasion by cells, especially fibroblasts, into tissue comprising an extracellular

matrix and/or migration by cells, especially fibroblasts, in or through tissue comprising an extracellular matrix.

In another embodiment, the present protein is used to prevent or reduce contracture of scar tissue resulting from eye surgery. Glaucoma surgery to create new drainage channels often fails
5 due to scarring and contraction of tissues. A method of preventing contraction of scar tissue formed in the eye, such as the application of a suitable agent, is therefore invaluable. Such an agent may also be used in the control of the contraction of scar tissue formed after corneal trauma or corneal surgery, for example laser or surgical treatment for myopia or refractive error in which contraction of tissues may lead to inaccurate results. It is also useful in cases where scar tissue is formed on/in
10 the vitreous humor or the retina, for example, that which eventually causes blindness in some diabetics and that which is formed after detachment surgery, called proliferative vitreoretinopathy. Other uses include where scar tissue is formed in the orbit or on eye and eyelid muscles after squint, orbital or eyelid surgery, or thyroid eye disease and where scarring of the conjunctiva occurs as may happen after glaucoma surgery or in cicatricial disease, inflammatory disease, for example,
15 pemphigoid, or infective disease, for example, trachoma. A further eye problem associated with the contraction of collagen-comprising tissues for which the methods and medicaments of the present invention may be used is the opacification and contracture of the lens capsule after cataract extraction.

In a preferred embodiment, the protein of the invention can be used for the treatment of
20 burns. Contraction of collagen-comprising tissue, which may also comprise other extracellular matrix components, frequently occurs in the healing of burns. The burns may be chemical, thermal or radiation burns and may be of the eye, the surface of the skin or the skin and the underlying tissues. It may also be the case that there are burns on internal tissues, for example, caused by radiation treatment.

25 A further aspect of the present invention is the inhibition of the contraction of skin grafts. Skin grafts may be applied for a variety of reasons and may often undergo contraction after application. As with the healing of burnt tissues the contraction may lead to both physical and cosmetic problems. It is a particularly serious problem where many skin grafts are needed as, for example, in a serious burns case.

30 An associated area in which the medicaments and methods of the present invention are of great use is in the production of artificial skin. To make a true artificial skin it is necessary to have an epidermis made of epithelial cells (keratinocytes) and a dermis made of collagen populated with fibroblasts. It is important to have both types of cells because they signal and stimulate each other using growth factors. A major problem up until now has been that the collagen component of the
35 artificial skin often contracts to less than one tenth of its original area when populated by fibroblasts. MMP inhibitors, for example, collagenase inhibitors may be used to inhibit the contraction to such an extent that the artificial skin can be maintained at a practical size.

Protein of SEQ ID NO:276 (157-15-4-0-B11-CS)

The protein of SEQ ID NO:276, encoded by the cDNA of SEQ ID NO:35, is a variant of a testis-specific isoform of human calpastatin protein (Genseq accession number W19395). The protein of SEQ ID NO:276 contains 2 potential transmembrane segments (position 5 to 25 and
 5 position 109 to 129) predicted by the software TopPred II (Claros and von Heijne, *CABIOS applic. Notes*, 10:685-686 (1994)), and a signal peptide (position 8 : LAVILTLLGLAIL/AI). Like the human calpastatin protein (Genseq accession number W19395), the protein of SEQ ID NO:276 is over-represented in testis.

Calpastatin is a physiological inhibitor of calpains. Calpains, a group of ubiquitous Ca^{2+} -
 10 activated cytosolic proteases, are thought to participate in cytoskeletal remodeling events, cellular adhesion, shape change, and mobility by the site-specific regulatory proteolysis of membrane- and actin-associated cytoskeletal proteins (Beckerle et al., *Cell* 51:569-577, 1987; Yao et al., *Am. J. Physiol.* 265(pt. 1):C36-46, 1993; and Shuster et al., *J. Cell Biol.* 128:837-848, 1995). Calpains have also been implicated in the pathophysiology of cerebral and myocardial ischemia, platelet
 15 activation, NF- κ B activation, Alzheimer's disease, muscular dystrophy, cataract progression and rheumatoid arthritis. There is considerable interest in inhibitors of calpain, as cellular adhesion, cytoskeletal remodeling events and cell mobility are linked to numerous pathologies (Wang et al., *Trends in Pharm. Sci.* 15:412-419, 1994; Mehdi, *Trends in Biochem. Sci.* 16:150-153, 1991). In addition, as the calpain/calpastatin system is involved in membrane fusion events for several cell
 20 types, and calpain can be detected in human sperm and testes extracts by Western blotting with specific antisera, tCAST may modulate calpain in the calcium-mediated acrosome reaction that is required for fertilization (Li S et al., *Biol Reprod*, 63(1):172-8, 2000).

Calpastatin consists of a unique N-terminal domain (domain L) and four repetitive protease-inhibitor domains (domains 1-4) (Lee WJ et al., *J Biol Chem*, 267(12):8437-42, 1992). The isolated
 25 cDNAs from various mammalian species have conspicuous differences in the regions encoding the N-terminal sequences and can be classified into four types. Alternative splicing is most likely the cause for the molecular diversity, and the multiple isoforms are implicated in specific physiological roles (Lee WJ et al., *J Biol Chem*, 267(12):8437-42, 1992). Type IV (or human tCAST), a shorter isoform, is specifically expressed in testis (Takano J et al., *J Biochem Tokyo*; 128(1):83-92, 2000).
 30 Human tCAST consists of a 40-amino-acid N-terminal T domain plus a part of domain II and all of domains III and IV from the somatic isoform. The protein of SEQ ID NO:276 shows extensive homology to the N-terminal region of the testis basic specific protein (U60665) and the human calpastatin protein (W19395). The homologous region corresponds to domain T and II of the human calpastatin protein (W19395). The T domain targets cytosolic localization and membrane
 35 association of tCAST, whereas domain I of somatic calpastatin proteins (sCAST) exhibits a nuclear localization function (Li S et al., *Biol Reprod*, 63(1):172-8, 2000).

It is believed that the protein of SEQ ID NO:276 is a member of the calpastatin family and, as such, plays a role in cytoskeletal remodeling events, cellular adhesion, shape change, and mobility by the site-specific regulatory proteolysis of membrane- and actin-associated cytoskeletal proteins. Preferred polypeptides of the invention are polypeptides comprising the amino acids of
5 SEQ ID NO:276 from positions 1 to 119. Other preferred polypeptides of the invention are any fragments of SEQ ID NO:276 having any of the biological activities described herein.

One embodiment of the present invention relates to methods of using the protein of the invention or part thereof in assays to detect the presence of calpain in a biological sample, such as in bodily fluids, in tissue samples, or in mammalian cell cultures. As calpastatin can bind calpain
10 (Murachi, *Biochemistry Int.*, 18(2)263-294, 1989), the protein of the invention can be used in assays and diagnostic kits to test the presence of calpain using techniques known to those skilled in the art. Preferably, a defined quantity of the protein of the invention or part thereof is added to the sample under conditions allowing the formation of a complex between the protein of the invention or part thereof, and the presence of the complex and/or the free protein of the invention or part thereof is
15 assayed and compared to a control. Calpastatin has been shown to be useful as a marker of intracellular calpain activation, and can be used for monitoring the involvement of calpain in pathological situations (De Tullio et al., *FEBS letter*, 475(1):17-21, 2000). Calpain has been implicated in cytoskeletal protein degradation involved in the pathophysiology of ischemia and disorders like Alzheimer's disease (Wronski et al., *J. Neural transm.*, 107(2):145-157, 2000),
20 apoptosis in neural cells of rat with spinal cord injury (SCI) (Ray, *Brain res.*, 867(1-2):80-9, 2000), cell fusibility (Kosower et al., *Methods Mol Biol.*, 144:181-94, 2000) and other physiopathologies. Assays detecting any increased calpain level in a cell would thus allow the diagnosis of any of the herein-described diseases or conditions. In addition, a recent study showed that in addition to their proteolytic activities on cytoskeletal proteins and other cellular regulatory proteins, calpain-
25 calpastatin systems can also affect expression levels of genes encoding structural or regulatory proteins (Chen et al., *Am. J. Physiol. Cell Physiol.*, 279:C709-C716, 2000). Thus, the ability to detect calpastatin and calpain levels will likely be useful for the diagnosis of an even larger number of diseases and conditions.

In another embodiment, the polynucleotides or polypeptides of the invention may be used
30 for the detection of gametes, or of specific structures within the gametes, using any technique known to those skilled in the art, including those involving the use of specific antibodies and nucleic acid probes. Various studies have shown that calpastatin is present in the sperm acrosome (Li et al., *Bio. of Reprod.*, 63:172-178, 2000), and more precisely between the plasma membrane and outer acrosomal membrane of cynomolgus macaque spermatozoa (Yudin AI, *J Androl*,
35 21(5):721-9, 2000). The ability to visualize spermatozoa generally, or the sperm acrosome in particular, has obvious utility for a number of applications, including for the analysis of infertility in patients, as described below.

Another embodiment of the present invention relates to a method of inhibiting a calpain in a cell. Various studies have shown that it is possible to inhibit calpains dose dependently in cell free protease activity assays: the calpain inhibitor Cerebrolysin can protect microtubule associated protein 2 (MAP2) in a rat model of acute brain ischemia (Wronski et al., J. Neural Transm. Suppl., 59:263-272, 2000), and E-64-D, a cell permeable and selective inhibitor of calpain, can attenuate calpain activity associated with apoptosis in rat SCI (Ray et al., Brain Res., 867(1-2)80-9, 2000). Similarly, it is believed that the protein of SEQ ID NO:276 can be used to inhibit calpain in vitro or in vivo. As calpain has been implicated in a number of pathological processes, diseases, and conditions, such as the pathophysiology of cerebral and myocardial ischemia, platelet activation, NF-kB activation, Alzheimer's disease, muscular dystrophy, cataract progression and rheumatoid arthritis, any of these diseases or conditions can be treated or prevented by increasing or decreasing the activity or expression of the present protein in cells of a mammal affected by the disease or condition. Such an increase can be effected in any of a number of ways, including, but not limited to introducing a polynucleotide encoding the protein of the invention, operably linked to a promoter, into a cell ; and by administering to a cell a compound that increases the activity or expression of the protein of the invention. In addition, the expression or activation of the protein of the invention can be inhibited in any of a large number of ways, including using antisense oligonucleotides, antibodies, dominant negative forms of the protein, and using heterologous compounds that decrease the expression or activation of the protein. Such compounds can be readily identified, e.g. by screening candidate compounds and detecting the level of expression or activity of the protein using any standard assay.

In another preferred embodiment, the protein of the invention can be used to modulate and/or characterize fertility, including for the treatment or diagnosis of infertility, and for contraception. As the calpain/calpastatin system has been implicated in the acrosomal reaction which is a required step in fertilization, it is likely that the over- or under-expression or activation of the present protein disrupts this reaction, thereby inhibiting fertility. Thus, the cause of infertility in many patients can likely be detected by detecting the level of expression of the present protein, where an abnormal level of activity or expression of the protein indicates that a cause of infertility involves the calpain-dependent acrosomal reaction. Such a diagnosis would also point to methods of treating the infertility, e.g. by increasing or decreasing the expression or activation of the protein in spermatozoa. Alternatively, for contraception, the expression or activation of the protein can be artificially disrupted, for example by increasing the protein level using polynucleotides encoding the protein, using the protein itself, or using activators of protein expression or activity, or by decreasing the protein level using inhibitors such as antisense oligonucleotides, antibodies, dominant negative forms of the protein, and using heterologous compounds that inhibit protein expression or activity.

Protein of SEQ ID NO: 295 (internal designation 181-20-3-0-B5-CS)

The protein of SEQ ID NO:295, encoded by the cDNA of SEQ ID NO:54, shows homology to the rat, bovine, and human uromodulin precursor, Tamm-Horsfall urinary glycoprotein, and human pancreatic secretory granule membrane major glycoprotein GP2 precursor. SEQ ID NO:295 exhibits
5 homology in the 5' region (over 40% identical and 60% similar) to both GP2 and uromodulin. Like GP2 and uromodulin, the homologous segment contains EGF-like calcium-binding domains, several potential disulfide bonds, and a number of potential N-linked glycosylation sites. Calcium binding EGF-like domains contain a calcium-binding site at the N-terminus, and have been found in proteins which require calcium for their biological activity. Non-limiting examples of proteins which contain
10 calcium-binding EGF-like domains include: (1) Coagulation Factors X, VII, IX; (2) LDL receptors; (3) thrombomodulin; and (4) fibrillin-1. Downing *et al.* [Cell 85:597-605 (1996)] described disease-causing mutations that destabilized a covalently-linked pair of Ca^{2+} -binding EGF-like domains in fibrillin-1 (associated with Marfan Syndrome). These domains form a rigid rod-like arrangement, stabilized by interdomain calcium binding and hydrophobic interaction. Uromodulin (URO) is a 90-
15 100 KDa glycoprotein synthesized by epithelial cells of the ascending loops of Henle and convoluted tubule of the bladder. Except for glycosylation, URO is identical to Tamm-Horsfall protein (THP), the most abundant protein in normal human urine. The relative abundance and specific nephronal location of URO suggests that it may have important physiologic functions in the urinary system.

URO has also been found to be an immunosuppressive glycoprotein, inhibiting antigen-
20 induced human T-cell proliferation. More recent studies have shown that URO can trigger the inflammatory response of neutrophils and stimulate human mononuclear cells to proliferate and release cytokines and gelatinase.

Uromodulin has been shown to play a role in regulating the circulating activity of cytokines since it binds to recombinant interleukin -1 and -2 and tumor necrosis factor (TNF) with high
25 affinity. Although URO does not inhibit the cytotoxic activity of $\text{TNF}\alpha$ as monitored by lysis of tumor cell targets, it interacts with recombinant $\text{TNF}\alpha$ via carbohydrate chains. This interaction may be critical in promoting clearance and/or reducing *in vivo* toxicity of $\text{TNF}\alpha$ and other lymphokines. Endotoxic shock and sepsis are caused by cytokines IL-1 and $\text{TNF}\alpha$. Since URO appears to exhibit inhibitory activity against IL-1 and $\text{TNF}\alpha$, URO may be effective as a therapeutic agent against
30 these conditions. Uromodulin has also been implicated as a possible inhibitor of certain types of bacterial infection in the bladder and urinary tract. URO has the ability to bind to type 1 pilus of *Escherichia coli* and prevent attachment to the surface of epithelium.

SEQ ID NO:295 also has homology to the glycoprotein GP-2. GP-2 is an integral protein of the pancreatic zymogen granule membrane. GP2 is anchored to the lipid bilayer via a glycosyl
35 phosphatidylinositol (GPI) linkage and released by a calcium-activated enzyme into the content of the zymogen granule. Through the process of exocytosis, GP2 is discharged into the pancreatic duct. The protein is also soluble in the zymogens stored in the granule, secreted by the pancreas,

and detected in the pancreatic secretions. GP2 appears to play a role in progression of pancreatitis, an inflammation of the pancreas accompanied by autodigestion of pancreatic tissue by its own enzymes. After cloning and sequencing of GP2, a search of the Genbank database revealed one homologous protein, namely uromodulin. Studies reveal that GP2 and URO not only share structural homology, but functionally are similar in that both can form ductal precipitates under pathological conditions. The aggregation of these precipitates in the pancreas may lead to obstruction of the pancreatic ducts and play a critical role in development of pancreatitis. Similarly, aggregation of URO in the kidney may lead to blockage of the renal tubules and result in renal disease.

10 The subject invention provides the protein of SEQ ID NO:295 and polynucleotide sequences encoding SEQ ID NO:295. Also included in the invention are biologically active fragments of the protein encoded by SEQ ID NO:295 and polynucleotide sequences encoding these biologically active fragments. "Biologically active fragments" are defined as those peptide or polypeptide fragments of SEQ ID NO:295 which have at least one of the biological functions of the full length protein (e.g., the ability to chelate calcium, bind to *E. coli* pili, or cause immunomodulation of an individual). In one embodiment, the polypeptides of SEQ ID NO:295 are interchanged with the polypeptides encoded by the human cDNA of clone 181-20-3-0-B5-CS.

The invention also provides variants of SEQ ID NO:295. These variants have at least about 80%, more preferably at least about 90%, and most preferably at least about 95% amino acid sequence identity to the amino acid sequence of SEQ ID NO:295. Variants according to the subject invention also have at least one functional or structural characteristic of SEQ ID NO:295, such as the biological functions described above or EGF-like calcium-binding domains. The invention also provides biologically active fragments of the variant proteins. Unless otherwise indicated, the methods disclosed herein can be practiced utilizing SEQ ID NO:295, or variants thereof. Likewise, the methods of the subject invention can be practiced using biological fragments of SEQ ID NO:295, or variants of said biologically active fragments.

Because of the redundancy of the genetic code, a variety of different DNA sequences can encode SEQ ID NO:295. It is well within the skill of a person trained in the art to create these alternative DNA sequences which encode proteins having the same, or essentially the same, amino acid sequence. These variant DNA sequences are, thus, within the scope of the subject invention. As used herein, reference to "essentially the same sequence" refers to sequences that have amino acid substitutions, deletions, additions, or insertions that do not materially affect biological activity.

"Recombinant nucleotide variants" are alternate polynucleotides which encode a particular protein. They can be synthesized, for example, by making use of the "redundancy" in the genetic code. Various codon substitutions, such as the silent changes which produce specific restriction sites or codon usage-specific mutations, can be introduced to optimize cloning into a plasmid or viral vector or expression in a particular prokaryotic or eukaryotic host system, respectively.

SEQ ID NO:295, and variants thereof, can be used to produce antibodies according to methods well known in the art. The antibodies can be monoclonal or polyclonal. Antibodies can also be synthesized against fragments SEQ ID NO:295 as well as variants of SEQ ID NO:295 according to known methods. The subject invention also provides antibodies which specifically
5 bind to biologically active fragments of SEQ ID NO:295 or biologically active fragments of variants of SEQ ID NO:295.

The subject invention also provides for immunoassays which are used to screen for, monitor, or diagnose conditions or disorders associated with liver dysfunction and/or damage. These conditions or disorders include, and are not limited to, hepatitis, cirrhosis, fibrosis,
10 pericholangitis, portal triaditis, chronic periportal inflammation, systemic lupus erythematosus, Hodgkin's disease, Granulomas, and cell dysplasia can also be diagnosed. For a number of disorders listed above, expression of these genes at significantly higher or lower levels can be routinely detected in certain liver tissues or cell types (e.g., cancerous) or bodily fluids (e.g., serum, plasma, and blood) taken from an individual having such a disorder, relative to the standard gene
15 expression levels, e.g., the expression level in healthy tissue or bodily fluid from one or more individuals not having the disorder. These types of assays allow for a non-invasive method of screening for, diagnosing, or monitoring liver cancer in human subjects. Similarly, antibodies and small molecules directed to the polypeptides can be used as immunological probes for differential identification of the diseased tissue(s) or cells.

20 Additionally, nucleic acid and amino acid sequences of SEQ ID NOs:54 and 295 can be used to provide polypeptides and biologically active fragments thereof for the repair of cellular injury following liver damage and/or liver transplant.

Furthermore, polypeptides, or biologically active fragments thereof, can be used for the modulation of bacterial binding to epithelial cells or as a modulator of bacterial infection. In this
25 aspect of the subject invention, bacterial cells are contacted with an amount of a composition comprising the polypeptide, or biologically active fragments thereof, sufficient to interfere with the binding of bacteria to epithelial cells. In one embodiment, the bacteria are coliform bacterial cells. In another embodiment, the bacterial cells are *E. coli*. Compositions comprising SEQ ID NO:295, or biologically active fragments thereof, can be administered in any fashion required to provide a
30 therapeutic effect (e.g., orally, intravenously, intrathecally, intraarterially, etc.).

The subject invention also provides materials and methods for the treatment of endotoxic shock and/or sepsis. In this embodiment, a subject can be treated with therapeutically effective amounts of a composition comprising SEQ ID NO:295, or biologically active fragments thereof.

The subject invention also provides materials and methods for the *in vivo* or *in vitro*
35 chelation of calcium ions (Ca^{2+}). In this aspect of the invention, SEQ ID NO: , or biologically active fragments thereof, can be used to bind free Ca^{2+} by addition of the polypeptide, or biologically active fragments thereof, to solutions, environmental samples, or biological samples.

Alternatively, a composition containing the SEQ ID NO:295, or biologically active fragments thereof, can be added to the solutions, environmental samples, or biological samples in amounts sufficient to bind and remove free Ca^{2+} from solution.

In another aspect of the subject invention, SEQ ID NO:295, or biologically active fragments thereof, can be used to modulate the immune system of a mammal. In this method, immunomodulatory amounts of SEQ ID NO:295, or biologically active fragments thereof, can be administered to a mammal in a pharmaceutically acceptable carrier. Methods of assessing the stimulated state of the immune system of the mammal can be practiced according to methods well known in the art.

10 Protein of SEQ ID NOs:244, 251 (internal designation numbers 105-016-3-0-G10-CS and 105-074-3-0-H10-CS)

The 274 amino acid protein of SEQ ID NO:244, encoded by the cDNA of SEQ ID NO:3, found in prostate and strongly expressed in the salivary gland, presents strong sequence similarities with the yeast putative mitochondrial carrier protein PET8 (SWISSPROT accession number P38921) and with similar proteins conserved among eukaryotes (*D. melanogaster* and *C. elegans*: respective SPTREMBLNEW SPTREMBL SWISSPROT accession numbers Q9VBN7 and Q18934, and *S. pombe*: SWISSPROT accession number: Q10442). All members of the mitochondrial carrier/transport protein superfamily exhibit sequence motifs highly similar to P-X-D/E-X-X-K/R-X-R that are also found in 3 positions in the protein of the invention (positions 26 to 33, 108 to 115 and 199 to 206) (Belenkiy *et al*, Biochim. Biophys. Acta, 1467:207-218 (2000)). These mitochondrial carrier protein signatures are associated with membrane-spanning segments (Belenkiy *et al*, *ibid*; Kuan et Saier, Crit. Rev. Biochem. Mol. Biol., 28:209-233 (1993)). In fact, 4 candidate membrane-spanning segments are identified in the protein of the invention, from amino acid positions 4 to 24, 51 to 71, 180 to 200 and 240 to 260. Other hydrophobic regions are found in positions 86 to 107 and 139 to 162. In addition, the protein of SEQ ID NO:244 presents a putative signal peptide in its very amino-terminal part (position 5 to 19).

The protein of SEQ ID NO:251, encoded by the cDNA of SEQ ID NO:10, is a 72 amino acid truncated form of the protein of SEQ ID NO:244. This shorter product results from the absence, in the cDNA of SEQ ID NO:10, of the 110bp exon (position 275 to 384) found in the cDNA of SEQ ID NO:3. Nevertheless, the 72 amino acid encoded protein possesses the putative signal peptide (position 5 to 19), the first mitochondrial carrier protein signature (position 26 to 33), and two candidate membrane-spanning segments (positions 4 to 24 and 51 to 71).

Energy transduction in mitochondria requires the transport of many specific metabolites across the inner membrane of this eukaryotic organelle. Different types of substrate carrier proteins involved in energy transfer are found in the inner membrane. These proteins all seem to be evolutionary related, and constitute the mitochondrial carrier/transport proteins (MCP/MTP)

superfamily. Structurally, MCP/MTP proteins are typically homodimeric integral transmembrane polypeptides (subunit molecular weight ~30kD) that traverse the inner mitochondrial membrane six times with both the N- and C-termini localized to the cytosolic side of the membrane. Each 30kD subunit is composed of three tandem repeats of a domain of approximately one hundred residues (~10kD). This 10kD domain contains two transmembrane regions and a sequence motif highly similar to P-X₁-D/E-X₂-X₃-K/R-X₄-R, where X₃ is a hydrophobic residue (Kuan et Saier, Crit. Rev. Biochem. Mol. Biol., 28:209-233 (1993)). Five protein families of known function have been identified among the mitochondrial carrier protein superfamily:

(1) The ADP, ATP carrier protein (ACC), ADP/ATP translocases, which under the conditions of oxidative phosphorylation catalyze the one to one exchange of cytosolic ADP against matrix ATP across the inner mitochondrial membrane (Fiore *et al*, Biochimie, 80:137-150 (1998)). The ADP/ATP transport system can be blocked very specifically by two families of inhibitors: atractyloside (ATR) and carboxyatractyloside (CATR) on one hand, and bongkrekic acid (BA) and isobongkrekic acid (isoBA) on the other hand. It is well established that these inhibitors recognise two different conformations of the carrier protein, the CATR- and BA-conformations, which exhibit different chemical, immunochemical and enzymatic reactivities. Bakker and collaborators have reported that myopathies might result from a defect in ADP/ATP transport (Bakker *et al*, Pediatr. Res. 33:412-417 (1993)). Namely, the authors describe a 4-fold decrease in the concentration of the ADP/ATP carrier protein in a patient with a mitochondrial myopathy.

(2) The 2-oxoglutarate/malate carrier protein (OGCP), which exports 2-oxoglutarate into the cytosol and imports malate, or other dicarboxylic acids, into the mitochondrial matrix. This protein plays an important role in several metabolic processes, such as the malate/aspartate and the oxoglutarate/isocitrate shuttles (Palmieri *et al*, J. Bioenerg. Biomembr. 25:493-501 (1993)).

(3) The phosphate group carrier protein, which transports inorganic phosphate groups from the cytosol into the mitochondrial matrix (Palmieri *et al*, *ibid*).

(4) The mammalian brown fat uncoupling proteins, such as UCP-1 (thermogenin), are transmembrane proton-translocating proteins present in the mitochondria of brown adipose tissue, a specialized tissue which functions in heat generation and energy balance ((Jezek and Garlid, Int. J. Biochem. Cell. Biol. 30:1163-1168 (1998); Klingenberg, J. Bioenerg. Biomembr. 31:419-430 (1999); Nicolls and Locke, Physiol. Rev. 64:2-40 (1994); Rothwell and Stock, Nature 281:31-35 (1979)). Mitochondrial oxidation of substrates is accompanied by proton transport out of the mitochondrial matrix, creating a transmembrane proton gradient. Typically, re-entry of protons into the matrix via ATP synthase is coupled to ATP synthesis. However, UCP-1 functions as a transmembrane proton transporter, permitting re-entry of protons into the mitochondrial matrix unaccompanied by ATP synthesis. Environmental exposure to cold evokes neural and hormonal stimulation of brown adipose tissue, which increases UCP mediated proton transport, brown fat metabolic activity, and heat production.

Studies with transgenic models indicate that brown fat and UCP-1 play an important role in energy expenditure in rodents. Transgenic mice in which brown adipocyte tissue was ablated by a toxin coupled to the UCP-promoter developed obesity and diabetes (Lowell *et al.*, Nature 366:740-742 (1993)). Obesity in these transgenic animals developed in the absence of hyperphagia, suggesting that the uncoupled mitochondrial respiration of brown fat is an important component of energy expenditure. In a separate transgenic mouse model, ectopic expression of UCP-1 in white adipose tissue of genetically-obese mice led to a significant reduction in body weight and fat stores (Kopecky *et al.*, J. Clin. Invest. 96:2914-2923 (1995)). These studies indicate that activity of UCP-1 is accompanied by energy expenditure and weight loss in rodents. Two other UCP proteins have recently been cloned. The first uncoupling protein-like protein (UCPL) or UCP-2 (59% homologous), is widely expressed (heart, kidney, lung, placenta and white fat) and enriched in tissues of the lymphoid lineage (Fleury *et al.*, Nature Genetics 15:269-272 (1997)). The second, UCP-3 (57% homologous), is predominantly localized to skeletal muscle and brown fat (Boss *et al.*, FEBS Lett. 408:39-42 (1997)). UCP-3 has been found to be regulated by cold and thyroid hormone (Larkins *et al.*, Biochem. Biophys. Res. Comm. 240:222-227 (1997)).

Thermogenic protein activity, such as that found with UCP-1, may be useful in reducing, or preventing the development of, excess adipose tissue, such as that found in obesity. Obesity is becoming increasingly prevalent in developed societies. Attempts to reduce food intake, or to decrease hypernutrition, are usually fruitless in the medium term because the weight loss induced by dieting results in both increased appetite and decreased energy expenditure (Leibel *et al.*, New Engl. J. Med. 322:621-628 (1995)). The intensity of physical exercise required to expend enough energy to materially lose adipose mass is too great for many obese people to undertake on a sufficiently frequent basis. Thus, obesity is currently a poorly treatable, chronic, essentially intractable metabolic disorder. In addition, obesity carries a serious risk of co-morbidities including, Type 2 diabetes, increased cardiac risk, hypertension, atherosclerosis, degenerative arthritis, and increased incidence of complications of surgery involving general anesthesia.

(5) The tricarboxylate transport protein (or citrate transport protein), which is involved in citrate-H⁺/malate exchange. This protein is important for the bioenergetics of hepatic cells as it provides a carbon source for fatty acid and sterol biosyntheses, and NAD for the glycolytic pathway (Kaplan *et al.*, J. Biol. Chem. 268:13682-13690 (1993)).

It is believed that the protein of SEQ ID NO:244 or part thereof is a member of the mitochondrial carrier/transport protein superfamily and, as such, plays a role in mitochondrial processes such as ADP/ATP, malate/aspartate, 2-oxoglutarate/isocitrate, citrate-H⁺/malate exchanges across the inner membrane, phosphate groups transport and physiological roles such as regulation of body weight and energy balance, muscle nonshivering thermogenesis, fever, and defense against the generation of reactive oxygen species. Preferred polypeptides of the invention are polypeptides comprising the amino acids of SEQ ID NO:244 from amino acid positions 26 to

33, 108 to 115 and 199 to 206 on one hand, and from positions 4 to 24, 51 to 71, 86 to 107, 139 to 162, 180 to 200 and 240 to 260 on the other hand. Other preferred polypeptides of the invention are fragments of SEQ ID NO:244 having any of the biological activities described herein. It is believed that the protein of SEQ ID NO:251 is a 72 amino acid truncated form of the 274 amino acid protein of SEQ ID NO:244, and corresponds to one subunit of the tripartite structure of mitochondrial carrier/transport proteins. Preferred polypeptides are polypeptides comprising the amino acids of SEQ ID NO:251 from positions 4 to 24, 26 to 33 and 51 to 71.

The activity of the protein of the invention can be assessed using cultured cells. For example, nucleic acids encoding the protein of SEQ ID NO:244 can be cloned into a eukaryotic vector and transfected into a population of cells. Transfected mammalian cells are then tested for their carrier activity e.g., the import of ADP, dicarboxylic acids, inorganic phosphate groups, or H^+ into the mitochondrial matrix, and the export of ATP, 2-oxoglutarate, tricarboxylate- H^+ export into the cytosol. These transfected cell lines may allow the development of *in vitro* assays for the identification of modulators of the carrier activity, such as atractyloside (ATR), carboxyatractyloside (CATR), bongkreikic acid (BA) and isobongkreikic acid (isoBA), which were described above in connection with the ADP/ATP mitochondrial carrier. Such modulators are useful for the treatment of any diseases or conditions associated with the protein of the invention.

Another embodiment of the invention relates to compositions and methods using the protein of the invention or part thereof to label mitochondria, or more specifically the inner mitochondrial membrane, in order to visualize any change in number, topology or morphology of this organelle, for example in association with a mitochondria-related human disorder, such as neuroleptic malignant syndrome (NMS) (Kubo et al., Forensic Sci. Int. 115:155-158 (2001)), the Rett syndrome (Armstrong, Brain Dev. 14 Suppl:S89-98 (1992)), Alpers disease (Chow and Thorburn, Hum. Reprod. 15 Suppl 2:68-78 (2000)) or mitochondrial encephalomyopathies (Handran et al., Neurobiol. Dis. 3:287-298 (1997)). For example, the protein may be rendered easily detectable by inserting the cDNA encoding the protein of the invention into a eukaryotic expression vector in frame with a sequence encoding a tag sequence. Eukaryotic cells expressing the tagged protein of the invention may also be used for the *in vitro* screening of drugs or genes capable of treating any mitochondria-related disease or conditions. The protein of the invention can also be used to specifically label cells of the salivary gland or of the prostate, e.g. for histological analyses or for the identification of the origin of tumor cells.

The protein of the invention can also be used as a carrier/transporter to translocate radiolabeled or chemically labeled metabolites (ADP, dicarboxylic acids, inorganic phosphate groups) from the cytosol to the matrix of the mitochondria in order to specifically label this organelle, e.g. to follow its modifications. For example, radiolabeled

or chemically labeled precursors can be added to an *in vitro* culture of mammalian cells stably transfected and expressing the protein of the invention. The labeling of the organelles can then be stopped at different times after the beginning of the experiment by adding specific inhibitors of carrier/transporter proteins, such as atractyloside (ATR),
5 carboxyatractyloside (CATR), bongkreikic acid (BA), or isobongkreikic acid (isoBA). Cells with labeled mitochondria can be used for the *in vitro* screening of drugs or genes capable of causing mitochondrial modifications.

Still another embodiment of the invention or part thereof relates to methods of delivering heterologous compounds, either polypeptides or polynucleotides, to the inner
10 membrane of mitochondria by recombinantly or chemically fusing a fragment of the protein of the invention to a heterologous polypeptide or polynucleotide. Preferred fragments are the putative peptide signal, the four membrane-spanning segments and/or any other fragments of the protein of the invention that may contain targeting signals for mitochondria including but not limited to matrix targeting signals as defined in Herrman
15 and Neupert, Curr. Opin. Microbiol. 3:210-4 (2000); Bhagwat et al. J. Biol. Chem. 274:24014-22 (1999), Murphy Trends Biotechnol. 15:326-30 (1997); Glaser et al. Plant Mol Biol 38:311-38 (1998); Ciminale et al. Oncogene 18:4505-14 (1999). Such heterologous compounds may be used to modulate mitochondrial activities, such as to induce and/or prevent mitochondrial-induced apoptosis or necrosis. For example, these
20 heterologous compounds may be used in the treatment and/or the prevention of disorders in which apoptosis is deleterious, including, but not limited to, immune deficiency syndromes (including AIDS), type I diabetes, pathogenic infections, cardiovascular and neurological injury, alopecia, aging, degenerative diseases such as Alzheimer's Disease, Parkinson's Disease, Huntington's disease, dystonia, Leber's hereditary optic neuropathy, schizophrenia,
25 and myodegenerative disorders such as "mitochondrial encephalopathy, lactic acidosis, and stroke" (MELAS), and "myoclonic epilepsy ragged red fiber syndrome" (MERRF). In addition, heterologous polynucleotides may be used to deliver nucleic acids for mitochondrial gene therapy, i.e. to replace a defective mitochondrial gene and/or to inhibit the deleterious expression of a mitochondrial gene.

30 The invention further relates to methods and compositions used to modify the protein of the invention. Post-translational modifications encompassed by the invention include, N-linked or O-linked carbohydrate chains, processing of N-terminal or C-terminal ends, attachment of chemical moieties to the amino acid backbone, chemical modifications of N-linked or O-linked carbohydrate chains, and addition or deletion of an N-terminal methionine residue as a result of prokaryotic host
35 cell expression. These post-translational modifications of the protein of the invention may be very

useful in the search for its putative protein partners, using approaches such as screening of an expression cDNA library with a radiolabeled recombinant protein, as post-translational modifications are of first importance in protein-protein interactions. Identification of proteinic partners of mitochondrial carrier proteins would allow the study of their regulation *ex vivo* and *in vivo* in normal versus pathologic cases (for an example concerning the UCP1 mitochondrial carrier protein and its 14.3.3 physical partner, see: Pierrat et al., Eur. J. Biochem. 267:2680-2687 (2000)).

Another embodiment of the invention relates to composition and methods using polynucleotide sequences encoding the protein of the invention or part thereof to establish transgenic model animals (*D. melanogaster*, *M. musculus*), by any method familiar to those skilled in the art. By modulating *in vivo* the expression of the transgene with drugs or modifier genes (activator or suppressor genes), animal models can be developed that mimic human mitochondria-associated disorders such as myopathies or obesity. These animal models would thus allow the identification of potential therapeutic agents for treatment of the disorders. In addition, recombinant cell lines derived from these transgenic animals may be used for similar approaches *ex vivo*.

In one embodiment, the protein of SEQ ID NO:251, corresponding to the 72 amino acid truncated form of SEQ ID NO:244, may be used as a dominant negative variant to inhibit the function of the full-length form of the protein of SEQ ID NO:244 *in vitro* or *in vivo*. Inactivation of mitochondrial carriers in this way may allow the development of animal models for human disorders. Recently, for example, Lowell and collaborators have shown in the mouse that a targeted destruction of UCP1 by the diphtheria toxin A chain is able to produce obese animals (Kozak and Koza, *ibid.*, Lowell et al., *ibid.*).

Protein of SEQ ID NO: 285 (internal designation 174-39-2-0-A3-CS)

The protein of SEQ ID NO:285, encoded by the cDNA of SEQ ID NO: 44 (clone 174-39-2-0-A3-CS), is overexpressed in cancerous prostate, fetal brain, muscle and placenta. The protein is homologous to the NADH-cytochrome b5 reductase isoform and to the human electron transport protein.

NADH-cytochrome b5 reductase proteins belong to a flavoenzyme family sharing common structural features and whose members (ferrodoxin-NADP⁺ reductase, NADPH-cytochrome P450 reductase, NADPH-sulfite reductase, NADH-cytochrome b5 reductase and NADH-nitrate reductase) are involved in photosynthesis, in the assimilation of nitrogen and sulfur, in fatty-acid oxidation, in the reduction of methemoglobin and in the metabolism of many pesticides, drugs and carcinogens (Karplus et al., Science, 251:60-6 (1991)). In addition, cytochrome b5 reductase is thought to play a role in the prevention of apoptosis following oxidative stress (see review by Villalba et al., Mol Aspects Med 18 Suppl:S7-13 (1997)).

It is believed that the protein of SEQ ID NO: 285 may be an oxidoreductase. Thus it may play a role in electron transport and general aerobic metabolism and may be associated with mitochondrial

membranes. In addition, the protein of the invention may be able to use FAD and/or molybdopterin as cofactors. It may be involved in photosynthesis, in the assimilation of nitrogen and sulfur, in fatty-acid oxidation, in the reduction of methemoglobin and in the metabolism of many pesticides, drugs and carcinogens. Preferred polypeptides of the invention are fragments of SEQ ID NO: 285 having any of the biological activity described herein. The oxidoreductase activity of the protein of the invention may be assayed using any technique known to those skilled in the art. The ability to bind a cofactor may also be assayed using any techniques well known to those skilled in the art including, for example, the assay for binding NAD described in US patent 5,986,172.

In another embodiment, the protein of the invention or part thereof is used to prevent cells from undergoing apoptosis. In a preferred embodiment, the apoptosis active polypeptide is added to an in vitro culture of mammalian cells in an amount effective to reduce apoptosis. Furthermore, the protein of the invention or part thereof may be useful in the diagnosis, the treatment and/or the prevention of disorders in which apoptosis is deleterious, including but not limited to immune deficiency syndromes (including AIDS), type I diabetes, pathogenic infections, cardiovascular and neurological injury, alopecia, aging, degenerative diseases such as Alzheimer's Disease, Parkinson's Disease, Huntington's disease, dystonia, Leber's hereditary optic neuropathy, schizophrenia, and myodegenerative disorders such as "mitochondrial encephalopathy, lactic acidosis, and stroke" (MELAS), and "myoclonic epilepsy ragged red fiber syndrome" (MERRF).

The invention further relates to methods and compositions using the protein of the invention or part thereof to diagnose, prevent and/or treat several disorders in which energy metabolism is impaired, or needs to be impaired, including but not limited to mitochondriocytopathies, necrosis, aging, neurodegenerative diseases, myopathies, methemoglobinemia, hyperlipidemia, obesity, cardiovascular disorders and cancer. For diagnostic purposes, the expression of the protein of the invention could be investigated using any of the Northern blotting, RT-PCR or immunoblotting methods described herein and compared to the expression in control individuals. For prevention and/or treatment purposes, the protein of the invention may be used to enhance electron transport and increase energy delivery using any of the gene therapy methods described herein.

Protein of SEQ ID NO:368 (internal designation 187-45-0-0-118-CS)

The protein of SEQ ID NO: 368 encoded by the cDNA of SEQ ID NO: 127 is a 78 amino acids long polypeptide. The sequence of the protein of SEQ ID NO: 368 is identical to the sequence of the human Dad1 protein, the defender against apoptotic cell death 1 protein, a subunit of the mammalian oligosaccharyltransferase (OST), except that the last 43 residues of Dad1 are replaced by a series of 8 different amino acids in the protein of the invention. In addition, the protein of SEQ ID NO: 368 displays the pfam signature for DAD family proteins from positions 1 to 78 as well as two putative transmembrane domains from positions 31 to 51 and 54 to 74. The Dad1 protein is a 113 amino acids long protein which mRNA is composed of 3 exons [see Genbank accession

number D15057 and Nakashima, T. *et al* (1993) *Molecular and Cellular Biology* 13:6367-6374].

The cDNA of SEQ ID NO: 127 is composed of the first and third exon of the Dad1 cDNA whereas the second exon of the Dad1 cDNA is missing. Taken together, these data indicate that the protein of SEQ ID NO: 127 is a new isoform of the Dad1 protein resulting from an alternative splicing

5 event.

Asparagine-linked glycosylation is a highly conserved protein modification reaction that occurs in all eukaryotes. The initial stage in the biosynthesis of N-linked glycoproteins, catalysed by the enzyme oligosaccharyltransferase (OST), involves the transfer of a preassembled high-mannose oligosaccharide from a dolichol-linked oligosaccharide donor onto asparagine acceptor sites in nascent proteins in the lumen of the rough endoplasmic reticulum [for review, see Silberstein, S. *et al* (1996) *FASEB J* 10: 849-858].

Protein glycosylation is essential for the structure and function of many proteins and is involved in the control of many diverse biological processes (Paulson, *Trends in Biol. Sci.*, 1989, 14, 272; Sadler, *In Biology of Carbohydrates*, 2nd Ed., Ginsburg & Robbins, Ed., John Wiley & Sons: New York, 1984, Vol. 2, pg. 87). For example, protein glycosylation has been found to be crucial for the development, growth and proper function of complex organisms, while the aberrant glycosylation of proteins has been associated with diseased and transformed cells.

The mammalian oligosaccharyltransferase is composed of the four ER membrane proteins, ribophorin I and II (RI and RII), OST48, and DAD1, which form an oligomeric complex. RI and OST48, and probably also RII, are type I transmembrane proteins. The luminal domain of OST48 interacts with those of RI and RII and the cytoplasmic domain of OST48 has affinity for the cytoplasmically exposed N-terminal tail of DAD1 [Kelleher, D. *et al.* (1997) *Proc Natl. Acad. Sci. USA* 94: 4994-4999; Fu, J. *et al.* (1997) *J. Biol. Chem* 272: 29687-29692].

Dad1 is a small hydrophobic protein, thought to be an integral membrane protein, with a cytoplasmically located N terminus and up to three transmembrane domains. As is true for the other subunits of OST, the precise role of Dad1 in N-glycosylation is not known. However, it has been shown that Dad1 is critical for the function and the structural integrity of the OST complex [Sanjay, A. *et al.* (1998) *J. Biol. Chem* 273: 26094-26099]. Also, it is worth noting that the Dad1 protein was first identified in 1993 as a mammalian cell death suppressor since loss of its function induces apoptosis in hamster BHK21 cells [Nakashima, T. *et al* (1993) *Molecular and Cellular Biology* 13:6367-6374]. Since then, several reports have confirmed the anti-apoptotic role for Dad1 [Hong, N.A. *et al.* (2000) *Dev Biol* 220:76-84; Brewster, J.L. *et al.* (2000) *Genesis* 26: 271-8; Yoshimi, M. *et al.* (2000) *Biochem Biophys Res Commun* 276: 965-9].

Dad1 is a highly conserved protein whose sequence has been determined for diverse organisms including several vertebrates, a nematode, and several plants. A comparison of these sequences reveals that the amino-terminal region preceding the first membrane-spanning segment is the least conserved region of the protein both with respect to length and amino acid sequence identity.

The most highly conserved sequences of Dad1 include the second and third membrane spanning segments, making them probably the most crucial regions for Dad1 function [Kelleher, D. *et al.* (1997) *Proc Natl. Acad. Sci. USA* 94: 4994-4999]. The importance of the C-terminus region for mediating Dad1 functions has been recently confirmed [Makishima, T. *et al.* (2000) *J. Biochem* 5 (Tokyo) 128:399-405]

Therefore, Dad1 is thought to act as a positive regulator of the oligosaccharyltransferase complex, and as a negative regulator of apoptosis. In addition, the C-terminus of the Dad1 protein seems to be important for mediating these functions. As mentioned above, the protein of the invention is a new isoform of the Dad1 protein resulting from an alternative splicing event. As a result of this alternative splicing event, the C-terminus of the protein of the invention is shortened and does not display the third transmembrane domain of Dad1. Since the C-terminus of Dad1 has been shown to be important for mediating the protein function, it is believed that the protein of the invention has rather an antagonistic action to the one of Dad1. It is worth noting that this type of situation in which the same gene give rise by alternative splicing to different protein products with opposing functions is a common theme among apoptosis genes [For a review, see Reed, JC. (1999) *Nat. Biotechnol* 17: 1064-65].

Thus, it is believed that the protein of the invention of SEQ ID NO: 368 plays a role in the control of N-glycosylation of cellular proteins. Preferably, the protein of the invention is thought to act as a positive regulator of apoptosis and a negative regulator of the OST complex. Preferred polypeptides of the invention are polypeptides comprising the amino acids of SEQ ID NO: 368 from positions 1 to 78, and 71 to 78. Other preferred polypeptides of the invention are fragments of SEQ ID NO: 368 having any of the biological activity described herein. The activity of the protein of the invention or part thereof on protein N-glycosylation may be assayed using any of the assays known to those skilled in the art. For example, one could use DNA-mediated gene transfer techniques in order to introduce the cDNA sequence of SEQ ID NO: 127 or part thereof into cell lines so that the protein of SEQ ID NO: 368 or part thereof is over expressed in these cell lines. The resulting effect of this over expression on the N-glycosylation of proteins can then be studied using immunoblotting or Western blotting of glycoproteins [Makishima *et al.* (1997) *Genes Cells* 2: 129-141; Silberstein *et al.* (1995) *J. Cell. Biol.* 131: 371-383; Hong *et al.* (2000) *Developmental Biology* 220: . The activity of the protein of the invention or part thereof on cellular apoptosis may be assayed using any of the assays known to those skilled in the art including those described by Nakashima *et al.* (1993) *supra*.

One object of the present invention are compositions and methods of targeting heterologous polypeptides to the endoplasmic reticulum by recombinantly or chemically fusing a fragment of the proteins of the invention to an heterologous polypeptide. Preferred fragments are any fragments of the proteins of the invention, or part thereof, that may contain targeting signals for the endoplasmic reticulum such as those described in Pidoux AL, Armstrong EMBO J 1992 Apr;11(4):1583-91; Munro

S, Pelham HR Cell 1987 Mar 13;48(5):899-907; Pelham HR Trends Biochem Sci 1990 Dec;15(12):483-6.

In another embodiment, the invention relates to compositions and methods using the protein of the invention or part thereof to stimulate cells' entry into apoptosis. In a preferred embodiment, the pro-apoptosis protein of the invention or part thereof is added to an in vitro culture of mammalian or plant cells in an amount effective to stimulate apoptosis. In another preferred embodiment, the cDNA sequence of SEQ ID NO: 127 or part thereof may be used to create transgenic animals or plant cells in which the disclosed protein of the invention or part thereof can be expressed at higher levels than normal whenever and wherever it is desired. Ways to create transgenic cells in which the expression of the transgene can be turned on or off whenever it is desired are well known in the art. Increasing the expression level of the protein of the invention in cells to stimulate programmed cell death may be useful for applications in which a given species of cells become undesirable upon a given event, i.e., infection, transformation, end of a production process, etc...

Furthermore, the invention relates to methods and compositions using the protein of the invention or part thereof to diagnose, prevent and/or treat disorders characterized by abnormal cell proliferation and/or programmed cell death, including but not limited to cancer, immune deficiency syndromes (including AIDS), type I diabetes, pathogenic infections, cardiovascular and neurological injury, alopecia, aging, degenerative diseases such as Alzheimer's Disease, Parkinson's Disease, Huntington's disease, dystonia, Leber's hereditary optic neuropathy, schizophrenia, and myodegenerative disorders such as "mitochondrial encephalopathy, lactic acidosis, and stroke" (MELAS), and "myoclonic epilepsy ragged red fiber syndrome" (MERRF). For diagnostic purposes, the expression of the protein of the invention could be investigated using any of the Northern blotting, RT-PCR or immunoblotting methods described herein and compared to the expression in control individuals. For prevention and/or treatment purposes of disorders in which cell proliferation needs to be reduced and/or apoptosis increased, the expression of protein of the invention may be enhanced using any of the gene therapy methods described herein or known to those skilled in the art. For prevention and/or treatment purposes of disorders in which cell proliferation needs to be enhanced and/or apoptosis reduced, inhibition of endogenous expression of the protein of the invention may be achieved using any methods or known to those skilled in the art including the triple helix and antisense strategies described herein.

Moreover, antibodies to the protein of the invention or part thereof may be used for detection of the endoplasmic reticulum for histological purposes using any techniques known to those skilled in the art.

Protein of SEQ ID No: 284 (internal designation 174-38-3-0-C9-CS)

The protein of SEQ ID No: 284 encoded by the cDNA of SEQ ID No: 43 is overexpressed in salivary gland. The 406-amino-acid-long protein of invention, which is similar in size to fucosyltransferases, displays a Pfam motif of the fucosyltransferase family from residues 70 to 406.

5 Furthermore, the present protein of invention is homologous to a putative fucosyltransferase of *Drosophila melanogaster* (STR accession number: Q9VLC1 and Q9VLC1). The protein of SEQ ID 284 also shares homology with the alpha1,3 fucosyltransferase (E.C. 2.4.1.152), found in *Brachydanio rerio* (EMBL accession number : AB023627), *Schistosoma mansoni* (GENPEPT accession number : AF183577-1), cattle (SPTREMBL accession number Q9TQQ3), and human
10 species (GENPEPT accession number : AJ132772_2). Like fucosyltransferases, the protein of the invention displays the features of type II transmembrane proteins with a short N-terminal cytoplasmic tail, a 9-29 amino-acid-signal-anchor transmembrane domain, and a large C-terminal domain. Furthermore, the present protein of invention displays an almost perfect consensus motif of the alpha-1,3 fucosyltransferases from residues 315 to 345 (Breton et al. Glycobiology 1998; 1: 87-
15 94).

Fucosyltransferases are a family of enzymes that catalyze the transfer of fucose from GDP-fucose, to galactose in an alpha1,2 linkage, and to N-acetylglucosamine in alpha1,3-, alpha1,4- and alpha1,6- linkages. Since all fucosyltransferases use the same nucleotide sugar, their specificity will probably reside in the recognition of the acceptor and in the type of linkage formed. In human
20 species, fucosyltransferases, which are type II membrane proteins found in Golgi, can be split into three distinct families (Breton et al. Glycobiology 1998; 1: 87-94): (1) the alpha-1,2-fucosyltransferases, hFUT1 and hFUT2, which yield nearly identical products as only single carbohydrate linkage differentiates type I from type II glycans. hFUT1 determines the expression of O-type antigen (H antigen) of the ABO blood group system on erythrocytes, whereas hFUT2 (Se)
25 determines it in saliva, *i.e.* secretor status; (2) The alpha-1,3-fucosyltransferases that constitute a distinct homogenous family of proteins, although some regions display similarities with the alpha-1,2 and alpha-1,6-fucosyltransferases (Breton et al. Glycobiology 1998;1:87-94). Five alpha -1,3-fucosyltransferases have been characterized to date in the human species, *i.e.* hFUT3 (Lewis enzyme), hFUT4 (myeloid-type), hFUT5, hFUT6 (plasma-type), and hFUT7. These are involved in
30 the last steps of the biosynthesis the carbohydrate antigen sialyl Lewis of ABH (de Vries et al. J Biol Chem 1995;270:8712-22 ; Kimura et al. Biochem Biophys Res Commun 1997 8;237:131-7) ; (3) The alpha-1,6-fucosyltransferase, hFUT8, which is implicated in the synthesis of N-glycans (Miyoshi et al. Biochim Biophys Acta 1999;1473:9-20).

The fucosylated cell surface glycoconjugates play important roles in physiological and
35 pathological processes, such as fertilization, embryogenesis, lymphocyte trafficking, immune response, and cancer metastasis (Staudacher et al. Trends Glycosci Glycotechnology 1996; 8:391-408). More specifically, the fucosylated cell surface glycoconjugates, which are present on the

apical surface of various epithelium, contribute to resistance of various microorganisms agents including bacteria as *Helicobacter pylori* (Umesaki et al. Science 1997;276:964-5), and *E. coli* (Vogeli et al. Schweiz Arch Tierheilkd 1997;139:479-84), and virus such as HIV (Ali et al. Infect Dis 2000;181:737-9). On the other hand, abnormal upregulation of fucosyltransferases is a common finding in various types of tumors, which cause an increased production of fucosylated glycoconjugates. Such fucosylated glycoconjugates can also serve as tumor markers and include (1) the Ca19-9 cancer antigen, which circulating sialyl-Lewis a structure produced by hFUT3 and used for diagnosis of pancreatic and gastric cancer (Koprowski et al. Somatic Cell Genet 1979;5:957-71), and (2) alpha-fetoprotein whose alpha 1,6-fucosylation is reduced in hepatoma (Miyoshi et al. Biochim Biophys Acta. 1999;1473:9-20). On the other hand, aberrant production of fucosylated glycoconjugates can provide selective growth advantage by facilitating the extravasation of tumor cells, since they participate to endothelial adhesion through interaction with E- and P-selectins of endothelial cells (Butcher and Picker, Science 1996;272:60-6). Consequently, modulation of fucosyltransferase activity can modify tumorigenicity in various model of tumors including hepatoma (Miyoshi et al. Biochim Biophys Acta. 1999;1473:9-20), and colorectal adenocarcinoma (Weston et al. Cancer Res 1999;59:2127-35).

Thus, it is believed that the protein of the invention of SEQ ID NO: 284 is a glycosyltransferase, preferably an hexosyltransferase, more preferably a fucosyltransferase, even more preferably an alpha-1,3-fucosyltransferase, and as such plays a role in fertilization, embryogenesis, lymphocyte trafficking, immune response, cancer metastasis and resistance to various microorganisms. Preferred polypeptides of the invention are polypeptides comprising the amino acids of SEQ ID NO: 284 from positions 70 to 406, and 315 to 345. Other preferred polypeptides of the invention are fragments of SEQ ID NO: 284 having any of the biological activity described herein. The glycosyltransferase activity of the protein of the invention or part thereof may be assayed using any of the assays known to those skilled in the art including those described in (Palcic et al. Carbohydr Res 1990;196:133-40).

Fucosylated compounds have considerable potential both as therapeutics and as reagents for clinical assays. However, synthesis of glycosylated compounds of potential commercial and/or therapeutic interest is difficult because of the very nature of the saccharide subunits. A multitude of positional isomers in which different substituent groups on the sugars become involved in bond formation, along with the potential formation of different anomeric forms, are possible. As a result of these problems, large scale chemical synthesis of most carbohydrates is not possible due to economic considerations arising from the poor yields of desired products. Enzymatic synthesis using glycosyl transferases such as fucosyltransferase provides an alternative to chemical synthesis of carbohydrates. Enzymatic synthesis using glycosidases, glycosyl transferases, or combinations thereof, have been considered as a possible approach to the synthesis of carbohydrates. As a matter of fact, enzyme-mediated catalytic synthesis would offer dramatic advantages over the classical

synthetic organic pathways, producing very high yields of carbohydrates economically, under mild conditions in aqueous solutions, and without generating notable amounts of undesired side products. To date, such enzymes are however difficult to isolate, especially from eukaryotic, e.g., mammalian sources, because these proteins are only found in low concentrations, and tend to be membrane-

5 bound. In addition to being difficult to isolate, the acceptor (peptide) specificity of glycosyl transferases is poorly understood. Thus, there is a need for obtaining recombinant glycosyl transferase, including fucosyltransferases, that could be produced in very large amounts.

Thus, the invention related to methods and compositions using the protein of the invention or part thereof to synthesize glycosylated compounds, either glycoproteins, glycolipids, or
10 oligosaccharides, more particularly fucosylated compounds. If necessary, the protein of the invention or part thereof may be produced in a soluble form by removing its transmembrane domains and/or its Golgi retention signal using any of the methods skilled in the art including those described in US patent 5,776,772. For example, the protein of the invention or part thereof is added to a sample containing GDP-fucose and a substrate compound in conditions allowing glycosylation,

15 more particularly fucosylation and allowed to catalyze the glycosylation of this compound. In a preferred embodiment, the enzymatic reaction carried out by the protein of the invention is part of a series of other chemical and/or enzymatic reactions aiming at the synthesis of complex glycosylated compounds, such as the ones described in US patents 5,409,817 and 5,374,541. In another preferred embodiment where the method is to be practiced on a commercial scale, it may be advantageous to
20 immobilize the glycosyltransferase on a support. This immobilization facilitates the removal of the enzyme from the batch of product and subsequent reuse of the enzyme. Immobilization of glycosyltransferases can be accomplished, for example, by removing from the transferase its membrane-binding domain, and attaching in its place a cellulose-binding domain. One of skill in the art will understand that other methods of immobilization could also be used and are described in the
25 available literature.

In a preferred embodiment, the present invention relates to processes and compositions for producing glycosylated compounds, preferably fucosylated compounds, wherein a cell is genetically engineered to produce the protein of the invention or part thereof and used in combination with one or several other cells able to produce the donor substrate for the protein of the
30 invention.

In another preferred embodiment, the present invention relates to a process and compositions for controlling the glycosylation of proteins in a cell wherein an insect, plant, or animal cell is genetically engineered to produce one or more enzymes that provide internal control of the cell's glycosylation mechanism. Preferably, the invention relates to a Chinese hamster ovary
35 (CHO) cell line that is genetically engineered to produce a fucosyltransferase of the present invention either alone or in combination with other glycosyltransferases. This supplemental fucosyltransferase modifies the glycosylation machinery to produce glycoproteins having

carbohydrate structures that more closely resemble naturally occurring human glycoproteins. The methods for performing the above process and making the above compositions are carried out using the methods known in the art and described in U.S. Patent No. 5,047,335.

Another embodiment of the present invention relates to compositions and methods using the protein or part thereof to detect fucosylated conjugates. In a preferred embodiment, the protein of SEQ ID No: 284 or part of thereof is used to obtain reagents, such as antibodies. These reagents could be used in radioimmunoassays, competitive binding assays, Western Blot analysis, enzyme-linked immunosorbant assay (ELISA), immunohistochemistry, or any other technique known to those skilled in the art (Palcic et al. Carbohydr Res 1990;196:133-40). In a preferred embodiment, antibodies raised against the present protein of invention provides tools to specifically visualize salivary or digestive tract tissues (and cells derived from the tissues). This can be useful for various applications, including the determination of the origin or identity of cells, e.g. cancerous cells, as well as to facilitate the identification of particular cells and tissues for, e.g. the evaluation of histological slides. Such assays may also be used for diagnosis in various disorders including, but are not limited to, neoplastic tumors such as salivary, prostate, liver, digestive disease tract and pancreas cancers. Various types of samples can be assayed, including tumor tissues, or other biological samples such as serum or plasma.

The invention further relates to glycosylated compounds, preferably fucosylated compounds, obtained using any of the processes described herein using the protein of the invention or part thereof. Such compounds may be used in the diagnosing, prevention and/or treating of disorders including, but are not limited to, cancer, cystic fibrosis, ulcer, inflammation and immune based disorders, including autoimmune disorders such as arthritis, fertility disorders, and hypothyroidism. These conditions include infectious diseases where active infection exists at any body site, such as meningitis and salpingitis; complications of infections including septic shock, disseminated intravascular coagulation, and/or adult respiratory distress syndrome; acute or chronic inflammation due to antigen, antibody and/or complement deposition; inflammatory conditions including arthritis, cholangitis, colitis, encephalitis, endocarditis, glomerulonephritis, hepatitis, myocarditis, pancreatitis, pericarditis, reperfusion injury and vasculitis. Immune-based diseases include but are not limited to conditions involving T-cells and/or macrophages such as acute and delayed hypersensitivity, graft rejection, and graft-versus-host disease; auto-immune diseases including type I diabetes mellitus and multiple sclerosis. In a preferred embodiment, these glycosylated compounds or derivatives thereof may be used as pharmacological agents to trap pathogens or endogenous ligands thus reducing the binding of pathogens or endogenous ligands to the endogenous glycosylated compounds. For example, such compounds may be used to prevent and/or inhibit the adhesion of cancer cells to inner wall of blood vessel or aggregation between cancer cells and platelets, thus reducing cancer metastasis, to prevent and/or inhibit the adhesion of neutrophils to blood vessels endothelial cells. Other disorders include infections in which

recognition of a glycosylated product is essential to the development of the infection. Such infections include, but are not limited to those caused by *Helicobacter pylori*, *E. coli* and viruses such as HIV. In a preferred embodiment, such compounds, preferably oligosaccharides, are used as gram positive antibiotics and disinfectants (U.S. Pat. Nos. 4,851,338 and 4,665,060).

5 The invention further relates to methods and compositions using the protein of the invention or part thereof for diagnosis, prevention and/or treatment of several disorders in which recognition of glycosylated compounds, preferably of fucosylated compounds, is impaired or needs to be impaired. For diagnostic purposes, the expression of the protein of the invention could be investigated using any of the Northern blotting, RT-PCR or immunoblotting methods described
10 herein and compared to the expression in control individuals. For prevention and/or treatment purposes, inhibiting the endogenous expression of the protein of the invention may be used to reduce the production of glycosylated compounds detrimental to the organism using any of the antisense or triple helix methods described herein as well as antagonists of the protein's activity.

 In another embodiment, various substances can be used for treatment, attenuation and/or
15 prevention for treatment of abnormal conditions associated to unbalanced amounts and/or activity of the protein of SEQ ID No. 284. Such substances include, but are not limited to, chemical compounds such as agonists and antagonists, nucleic acids, and antibodies. In particular, the protein of the invention or part thereof may be used in the development of inhibitors of glycosyl transferase, more particularly inhibitors of fucosyltransferases, for mechanistic and clinical applications (Taylor,
20 Curr Opin Struc Biol 1996;6:830-7 ; Colman, Pure Appl Chem 1995;67:1683-8; Bamford, Enz Inhib 1995;10:1-16 ; Khan & Matta, *In Glycoconjugates, Composition, Structure, and Function*. pp361-378. eds., Allen, H. J. & Kisailus, E. C. Marcel Dekker, Inc. New York, 1992 ; Thorne-Tjomsland et al., Transplantation 2000;69:806-8 ; Basset et al., Scand J Immunol 2000;51:307-11). Such substances may be employed for treatment of a variety of therapeutic and prophylactic
25 purposes including certain types of neoplastic disorders. For instance, substances targeted against the protein of SEQ ID No. 284 can be administered to treat patients affected with, but not limited to, salivary, prostate, liver, digestive disease tract and pancreas cancers. Alternatively, such substances can be used for treatment, attenuation and/or prevention of infectious disease in order to induce resistance of various microorganisms agents.

30 Protein of SEQ ID NO:292 (internal designation 181-10-1-0-D10-CS)

 The protein of SEQ ID NO:292 is encoded by the cDNA of SEQ ID NO:51. Accordingly, it will be appreciated that all characteristics and uses of the polypeptide of SEQ ID NO:292 described throughout the present application also pertain to the polypeptide encoded by a nucleic acid included in clone 181-10-1-0-D10-CS. In addition, it will be appreciated that all characteristics
35 and uses of the nucleic acid of SEQ ID NO:51 described throughout the present application also pertain to the nucleic acid included in clone 181-10-1-0-D10-CS.

The protein of SEQ ID NO:292 was identified among the cDNAs from a library constructed from fetal liver. Tissue distribution analysis using databases indicated that mRNA encoding this protein was found primarily in fetal kidney and fetal liver.

The protein of SEQ ID NO:292 is most likely a polymorphic variant (92% identity) of human secreted protein SEQ ID NO: 197 from the protein described in PCT publication WO 9906553-A2, the disclosure of which is incorporated herein by reference in its entirety. Further, the protein of SEQ ID NO:292 is homologous to the C-type lectin domain of mouse macrophage asialoglycoprotein-binding protein (M-ASGP-BP, 36% identity), mouse natural killer (NK) cell surface protein P1 40 (NKR-, P1.9, 34%) and human asialoglycoprotein receptor L-H2 from EP 773289-A2 (27%). Thus, the present invention relates to nucleic acid and amino acid sequences of a lectin-like protein and to the use of these sequences in the diagnosis, study, prevention and treatment of disease.

The protein of SEQ ID NO:292 consists of 111 amino acids. From the amino acid alignments and the hydrophobicity plots, it has a predicted signal peptide sequence spanning residues 12-24 and one predicted transmembrane domain spanning residues 5-25. Accordingly, one embodiment of the present invention is a polypeptide comprising the signal peptide or the transmembrane domain.

A number of different protein families share a conserved domain which was first characterized in some animal lectins and which seems to function as a calcium-dependent carbohydrate-recognition domain (Drickamer K., J. Biol. Chem., 263:9557-9560, 1988, the disclosure of which is incorporated herein by reference in its entirety). This domain, which is known as the C-type lectin domain (CTL) or as the carbohydrate-recognition domain (CRD), consists of about 110-130 residues. There are four cysteines that are perfectly conserved and involved in two disulfide bonds. Several categories of proteins can be found in which the CTL domain has been described. Both M-ASGP-BP and NKR-P1 are type II membrane proteins. Type II membrane proteins in which the CTL domain has been located at the C-terminal extremity include: 1) Asialoglycoprotein receptors (ASGPR), also known as hepatic lectins. The ASGPR's mediate the endocytosis of plasma glycoproteins to which the C-terminal sialic acid residue in their carbohydrate moieties has been removed. 2) A number of proteins expressed on the surface of NK cells, and some subsets of T cells: NKG2, NKR-P1, Ly-49, CD69, and on B cells: CD72, LyB-2. The CTL- domain in these proteins is distantly related to other CTL-domains, and it is unclear whether they all bind carbohydrates.

M-ASGP-BP is a lectin-like molecule expressed on the surface of activated macrophages and specific for terminal D-galactose and N-acetyl-D-galactosamine units (Oda S et al., J. Biochem., 104:600-605, 1988, the disclosure of which is incorporated herein by reference in its entirety). Experimental results suggest that M-ASGP-BP participates in the interaction between tumoricidal macrophages and tumor cells.

ASGPR is a membrane protein expressed specifically by hepatocytes. Its function is to uptake asialoglycoproteins in the serum for degradation in the liver. Partially deglycosylated plasma glycoproteins are efficiently and specifically removed from the circulation by a receptor-mediated process. In mammals, the ASGPR specific for desialylated (galactosyl-terminal) glycoproteins, is expressed exclusively in hepatic parenchymal cells. Following binding of the ligand to this cell surface receptor, the receptor-ligand complex is internalized and transported by a series of membrane vesicles and tubules to an acidic-sorting organelle where receptor and ligand dissociate (Spiess M et al., J. Biol. Chem., 260:1979-1982, 1985, the disclosure of which is incorporated herein by reference in its entirety). Reduction in expression of AGPR has been reported in response to such liver conditions as hepatic cirrhosis, liver cancer and regenerated liver (Stadalnik et al., J. Nucl. Med., 26:1233-1242, 1985, the disclosure of which is incorporated herein by reference in its entirety). It has also been reported that ASGPR itself is present in serum (Katsugi et al., Alcohol Metabolism and Liver, 12:65-68, 1992, the disclosure of which is incorporated herein by reference in its entirety), which resulted in significant research being pursued toward the measurement of serum ASGPR. Furthermore, published results indicate that labeling compounds binding to ASGPR can be used as good indicators of liver function (Kudo, et al., Japan Assoc. of Gastrointest. Pathology., 89:1349-1359, 1992, the disclosure of which is incorporated herein by reference in its entirety).

NK cells constitute the third major population of lymphocytes. They possess the inherent capacity to kill various tumors and virally infected cells and mediate the rejection of allografts. These properties allow NK cells to have a major role in the regulation of innate immune responses in particular, and immunological functions in general. Members of the NKR-P1 family are type-II transmembrane C-type lectin receptors found on the surface of NK cells and a subset of T lymphocytes (NK T cells). Further, a subset of NKR-P1 molecules has been identified at the surface of peripheral blood monocytes and dendritic cells (Poggi A et al., Eur. J. Immunol., 27: 2965-2970, 1997, the disclosure of which is incorporated herein by reference in its entirety). Deficiencies in NKR-P1⁺ T cells, which preferentially accumulate in the liver and bone marrow, have been implicated in the susceptibility to many diseases including insulin-dependent diabetes mellitus (IDDM, Tori M et al. Transplantation, 70:32-38, 2000, the disclosure of which is incorporated herein by reference in its entirety) and multiple sclerosis (Poggi A et al., J. Immunol., 162: 4349-4354, 1999, the disclosure of which is incorporated herein by reference in its entirety). NKR-P1 receptors have been shown to activate NK cell cytotoxicity coupled with release of interferon- γ (IFN- γ , Brown MG., Immunol. Rev., 155: 55-75, 1997, the disclosure of which is incorporated herein by reference in its entirety). However, unlike the well-characterized MHC class I ligands that regulate the specificity of the Ly-49 family of molecules, which are structurally related to the NKR-P1 receptors, cognate ligands for the NKR-P1 molecules have yet to be identified. Interestingly, it has been reported that a subset of the NKR-P1 molecules- NKR-P1B -

inhibits NK cell activation (Carlyle et al., J. Immunol., 162:5917-5923, 1999, the disclosure of which is incorporated herein by reference in its entirety).

Based on the structural and chemical homologies the protein of SEQ ID NO:292 was characterized as a C-type lectin-like, type II membrane protein, whose ligand binding may be calcium dependent. The protein of SEQ ID NO:292 or fragments thereof may provide the basis for clinical diagnosis of diseases associated with its induction and/or repression. This protein, fragments thereof or antagonists/inhibitors thereof may be useful in the diagnosis and treatment of tumors, viral infections, inflammation, or conditions associated with impaired immunity, organ transplantation, bacterial infections, autoimmunity, hepatic dysfunction and liver regeneration. Furthermore, the protein SEQ ID NO:292 or fragments thereof may be used as a reagent for analyzing the control of gene expression by IFNs and other cytokines such as IL-12 and IL-4, as well as growth and transcription factors, in normal and diseased cells.

The protein of SEQ ID NO:292 has homology to the CTL domains of the ASPRG, M-ASGP-BP and NKR-P1 molecules. The protein of SEQ ID NO:292, in membrane-bound or soluble forms may have cytokine receptor activity, cell proliferation/differentiation activity, T cell activation activity, tissue growth regulating activity, receptor/ligand activity, signal transduction activity, to promote transendothelial migration, anti-inflammatory activity, tumor inhibition activity, among others. Accordingly, the protein SEQ ID NO:292 or fragments thereof may be used in diagnosis and treatment of diseases such as, but not limited to, autoimmune disorders such as autoimmune hepatitis, rheumatoid arthritis, Graves disease, systemic lupus erythematosus, Wegener's granulomatosis, sarcoidosis, polyarthritis, pemphigus, pemphigoid, erythema multiform, Sjogren's syndrome, inflammatory bowel disease, autoimmune encephalitis, myasthenia gravis keratitis, scleritis, Lupus Nephritis, and allergic encephalomyelitis; proliferative disorders including various forms of cancer such as leukemias, lymphomas (Hodgkins and non-Hodgkins), sarcomas, melanomas, adenomas, carcinomas of solid tissue, hypoxic tumors, squamous cell carcinomas of the mouth, throat, larynx, and lung, genitourinary cancers such as cervical and bladder cancer, hematopoietic cancers, head and neck cancers, and nervous system cancers, benign lesions such as papillomas, atherosclerosis, angiogenesis; viral infections, in particular HBV, HCV and HIV infections, as well as other viral- and pathogen-induced infections. The protein of SEQ ID NO:292 or fragments thereof may also be used to treat conditions associated with inflammation or immune impairment (e. g. reumathoid and osteo arthritis and AIDS), allergy, hepatic cirrhosis and liver toxicity; as well as genetic disorders, chronic illnesses and infections associated with decrease in NK, NK T, macrophage, monocyte and dendritic cell functions. In another embodiment of the invention, inhibitors of the protein of SEQ ID NO:292 may be used to treat conditions such as multiple sclerosis, IDMM, graft versus host disease (GVH) and transplanted organ rejection.

Another embodiment relates to methods to treat and/or prevent the bacterial infections that arise in liver due to bacterial antigens brought from the intestine from the portal vein. In this

embodiment, the protein of SEQ ID NO:292 may be used to counteract the effects of the bacterial endotoxin lipopolysaccharide (LPS). Another embodiment of the invention is the use of the protein of SEQ ID NO:292 or fragments thereof to inhibit of NK cells activated by bacterial superantigens or LPS, which would help treat vascular endothelial injury in conditions such as Kwasaki disease.

5 The appearance of autoantibodies against the protein of SEQ ID NO:292 can be used as an indicator for autoimmune hepatitis (AIH), a disease that can lead to cirrhosis and fatal intractable hepatitis, as well as primary biliary cirrosis. The nucleic acid sequences encoding the protein of SEQ ID NO:292 or fragments thereof can be used for producing secreted forms of the protein. They can also be used to develop products for diagnosis and therapy. Accordingly, recombinant soluble
10 derivatives can be used for detecting and measuring antibodies specific for the protein of the invention, e.g. by ELISA, Western blotting, etc. This allows AIH to be diagnosed and distinguished from other diseases.

In another embodiment of the invention, the protein of SEQ ID NO:292 or fragments or derivatives thereof can also be used for the analysis and purification of asialoglycoproteins and to
15 develop inhibiting agents against asialoglycoprotein incorporation, or viral and other protein invasion, into liver cells.

Another embodiment of the present invention relates to polypeptides comprising the protein of SEQ ID NO:292 or fragments thereof and polynucleotides encoding the protein of SEQ ID NO:292 or fragments thereof. In another aspect the protein of SEQ ID NO:292 or fragments thereof
20 may be used to identify specific molecules with which it binds such as agonists, antagonists or inhibitors. In a further aspect, the invention relates to methods for identifying agonists and antagonists/inhibitors of the protein of SEQ ID NO:292, and treating conditions associated with the protein of the invention or imbalance with the identified compounds. In a still further aspect, the invention relates to diagnostic assays for detecting diseases associated with inappropriate levels or
25 activity of the protein of SEQ ID NO:292. Another embodiment of the invention relates to methods of measuring the amount of the protein of SEQ ID NO:292 in serum. Another embodiment relates to the use of labeling compounds that bind to the protein of SEQ ID NO:292 and can be used as good indicators of liver function or NK cell activity, among others.

An embodiment of the present invention relates to methods of using the protein of the
30 invention or part thereof to identify and/or quantify or other ligands, which may interact with the protein of SEQ ID NO:292. The protein of SEQ ID NO:292 or fragments thereof may be include in pharmaceutical preparations for treating cancer or prevention/treatment of other diseases associated with changes in expression of the protein of the invention (see above). In a preferred embodiment of the invention the protein of the invention or part thereof is used to modulate the effect of
35 cytokines and related molecule such as IL-1, IL-2, IL-12, IFN- γ . The protein of SEQ ID NO:292 may also be used to correct defects in vivo models of disease such as autoimmune, inflammation, pathogen-mediated infection, liver toxicity, allograft rejection, GVH, as well as tumor models, by

injecting the protein either intraperitoneally, intravenously, subcutaneously or directly in the diseased tissue.

The DNA encoding the protein of SEQ ID NO:292 or fragments thereof may be used in diagnostic assays for conditions/diseases associated with up-regulation or down-regulation of the expression of the protein of the invention (see above). The diagnostic assay is useful to distinguish between absence, presence, and excess expression of the protein and to monitor regulation of levels of the protein of during therapeutic intervention. The DNA may also be incorporated into effective eukaryotic expression vectors and directly targeted to a specific tissue, organ, or cell population for use in gene therapy to treat the above mentioned conditions, including tumors and/or to correct disease- or genetic-induced defects in any of the above mentioned proteins including the protein of the invention. The DNA may also be used to design antisense sequences and ribozymes, which can be administered to modify gene expression in NK, NK T, macrophages, monocytes and dendritic cells and to influence expression of cytokines such as IL-1, IL-2, IL-4, IL-12, and IFN- γ . In vivo delivery of genetic constructs into subjects can be developed to the point of targeting specific cell types, such as tumor where expression of the protein of SEQ ID NO:292 may be affected or is modulating the expression and/or activity of other proteins such as cytokines, growth factors, their receptors and/or tumor antigens. The DNA may also be used to identify unknown upstream sequences (e. g. promoters and regulatory elements) by standard techniques and for research into the control of gene expression by IFNs and other cytokines, as well as growth and transcription factors in normal and diseased cells. Hybridization probes are useful to detect DNA encoding the protein of SEQ ID NO:292 (or closely related molecules) in biological samples, and for mapping the naturally occurring genomic sequence to a particular chromosome/chromosome region. The DNA may be used to generate and/or treat in vivo animal models of disease, including susceptibility or resistance to infection, tumors, autoimmune conditions, GVH, allograft rejection and liver toxicity, based on vaccine, knock-out and transgene technologies.

Antibodies against the protein of SEQ ID NO:292 are useful for the diagnosis of conditions and disease associated with its expression and to quantify the protein of the invention (e. g. in assays to monitor patients during therapeutic intervention). Antibodies specific for the protein may include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, Fab fragments produced by a Fab expression library. Neutralizing antibodies are especially preferred for diagnostics and therapeutics. Diagnostic assays for the protein of the invention include methods utilizing the antibody and a label to detect the protein of SEQ ID NO:292 in human body fluids or extracts of cells or tissues as well as methods for detecting or measuring antibodies against the protein of SEQ ID NO:292.

The protein of SEQ ID NO:292 and its catalytic or immunogenic fragments or oligopeptides thereof, can be used for screening therapeutic compounds in any variety of drug screening techniques including high throughput. Methods which may be used to quantitate the expression of

the nucleotide or protein of the invention include, but are not limited to, polymerase chain reaction (PCR), RT-PCR, RNase protection, Northern blotting, enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), fluorescent activated cell sorting (FACS), immunoprecipitation, and chromatography.

5 Accordingly, the protein of SEQ ID NO:292 or fragments thereof may be used to purify or enrich proteins containing carbohydrates. In such embodiments, the lectin of the present invention is placed in contact with carbohydrate-containing proteins under conditions which facilitate specific binding. The lectin of the present invention may be fixed to a solid support. After binding, specifically bound proteins are dissociated using appropriate salt or other conditions.

10 The protein of SEQ ID NO:292 or fragments thereof may also be used to regulate any of the activities described above, including the interaction between tumoricidal macrophages and tumor cells, the activity of NK cells, the treatment of bacterial infections resulting from bacterial antigens brought from the intestine, or to counteract the effects of bacterial LPS.

 Accordingly, the present invention includes the use of the protein of SEQ ID NO:292 ,
15 fragments comprising at least 5, 8, 10, 12, 15, 20, 25, 30, 35, 40, 50, 60, 75, 100, 150, or 200 consecutive amino acids thereof, or fragments having a desired biological activity to treat or ameliorate a condition in an individual. For example, the condition may be any of those described above or an abnormality in any of the functions listed above. In such embodiments, the protein of SEQ ID NO:292, or a fragment thereof, is administered to an individual in whom it is desired to
20 increase or decrease any of the activities of the protein of SEQ ID NO:292. The protein of SEQ ID NO:292 or fragment thereof may be administered directly to the individual or, alternatively, a nucleic acid encoding the protein of SEQ ID NO:292 or a fragment thereof may be administered to the individual. Alternatively, an agent which increases the activity of the protein of SEQ ID NO:292 may be administered to the individual. Such agents may be identified by contacting the
25 protein of SEQ ID NO:292 or a cell or preparation containing the protein of SEQ ID NO:292 with a test agent and assaying whether the test agent increases the activity of the protein. For example, the test agent may be a chemical compound or a polypeptide or peptide.

 Alternatively, the activity of the protein of SEQ ID NO:292 may be decreased by administering an agent which interferes with such activity to an individual. Agents which interfere
30 with the activity of the protein of SEQ ID NO:292 may be identified by contacting the protein of SEQ ID NO:292 or a cell or preparation containing the protein of SEQ ID NO:292 with a test agent and assaying whether the test agent decreases the activity of the protein. For example, the agent may be a chemical compound, a polypeptide or peptide, an antibody, or a nucleic acid such as an antisense nucleic acid or a triple helix-forming nucleic acid.

35 In one embodiment, the invention relates to methods and compositions using the protein of the invention or part thereof as a marker protein to selectively identify the source of a sample as, for example, fetal liver or fetal kidney, or to distinguish between two or more possible sources of a

sample on the basis of the level of the protein of SEQ ID NO:292 in the sample. For example, the protein of SEQ ID NO:292 or fragments thereof may be used to generate antibodies using any techniques known to those skilled in the art, including those described therein. Such antibodies may then be used to identify tissues of unknown origin, for example, forensic samples, differentiated tumor tissue that has metastasized to foreign bodily sites, or to differentiate different tissue types in a tissue cross-section using immunochemistry. In such methods a sample is contacted with the antibody, which may be detectably labeled, under conditions which facilitate antibody binding. The level of antibody binding to the test sample is measured and compared to the level of binding to control cells from fetal liver or fetal kidney or tissues other than fetal liver or fetal kidney to determine whether the test sample is from fetal liver or fetal kidney. Alternatively, the level of the protein of SEQ ID NO:292 in a test sample may be measured by determining the level of RNA encoding the protein of SEQ ID NO:292 in the test sample. RNA levels may be measured using nucleic acid arrays or using techniques such as in situ hybridization, Northern blots, dot blots or other techniques familiar to those skilled in the art. If desired, an amplification reaction, such as a PCR reaction, may be performed on the nucleic acid sample prior to analysis. The level of RNA in the test sample is compared to RNA levels in control cells from fetal liver or fetal kidney or tissues other than fetal liver or fetal kidney to determine whether the test sample is from fetal liver or fetal kidney.

In another embodiment, antibodies to the protein of the invention or part thereof may be used for detection, enrichment, or purification of cells expressing the protein of SEQ ID NO:292, including using methods known to those skilled in the art. For example, an antibody against the protein of SEQ ID NO:292 or a fragment thereof may be fixed to a solid support, such as a chromatography matrix. A preparation containing cells expressing the protein of SEQ ID NO:292 is placed in contact with the antibody under conditions which facilitate binding to the antibody. The support is washed and then the cells are released from the support by contacting the support with agents which cause the cells to dissociate from the antibody.

In another embodiment of the present invention, the protein of SEQ ID NO:292 or a fragment thereof thereof may be used to diagnose disorders associated with altered expression of the protein of SEQ ID NO:292. In such techniques, the level of the protein of SEQ ID NO:292 in an ill individual is measured using techniques such as those described herein. The level of the protein of SEQ ID NO:292 in the ill individual is compared to the level in normal individuals to determine whether the individual has a level of the protein of SEQ ID NO: 292 which is associated with disease.

Protein of SEQ ID NO: 408 (internal designation 179-14-2-0-F11-CS)

The 236 amino acid protein of SEQ ID NO: 409, herein referred to as PNMT A, and encoded by the cDNA of SEQ ID NO: 168 is found in fetal kidney and fetal brain. PNMT A is a

polymorphic variant of human phosphatidylethanolamine N-methyltransferase (PNMT) (SPTREMBLNEW SPTREMBL SWISSPROT accession number Q9UHY6). PNMT A differs from the sequence of PNMT (STR accession number Q9UHY6) by two amino residues. Position 95 contains an isoleucine residue (I) substituted for a valine residue (V); position 130 contains a valine residue (V) substituted for a glutamine residue (G). PNMT A displays 4 candidate membrane-spanning segments in positions 50 to 70, 83 to 103, 131 to 151 and 196 to 216.

Catecholamine neurotransmitters [e.g., dopamine, noradrenaline (norepinephrine), adrenaline (epinephrine)] are synthesized in catecholaminergic neurons from tyrosine, via dopa, dopamine and noradrenaline, to adrenaline. Four enzymes are involved in the biosynthesis of adrenaline: (1) tyrosine 3-mono-oxygenase (tyrosine hydroxylase, TH); (2) aromatic L-amino acid decarboxylase (AADC, or Dopa decarboxylase, DCC); (3) dopamine beta-mono-oxygenase (dopamine beta-hydroxylase, DBH); and (4) noradrenaline N-methyltransferase (phenylethanolamine N-methyltransferase, PNMT) (Nagatsu, *Neurosci. Res.* 12:315-345 (1991)). PNMT, the final enzyme in the pathway for adrenaline biosynthesis catalyses the production of adrenaline from noradrenaline using S-adenosyl-L-methionine as a methyl donor. For this reason, PNMT serves as a good marker for tissues and cells producing epinephrine (adrenaline). Studies conducted by Kennedy and collaborators have shown that PNMT are widely distributed in human tissues including heart and kidney (Kennedy et al., *J. Clin. Invest.* 95:2896-2902 (1995)).

In some pheochromocytomas, the tumors contain and secrete greater amounts of adrenaline than do normal adrenal medullas. In a case/control study, Isobe *et al.* have shown that adrenaline-secreting pheochromocytomas express significantly greater amounts of PNMT mRNA than do normal adrenal medullas (Isobe *et al.* *J. Urol.* 163:357-362 (2000)). Moreover, PNMT immunoreactivity is only detected in the adrenaline-secreting tumors. The C-1 region in the rostral ventral lateral medulla contains mainly adrenaline neurons. These neurons are the tonic vasomotor center of the brain. Burke *et al.* have demonstrated changes in the enzymatic activity of PNMT in axon terminals and cell bodies of neurons from the medulla of patients with Alzheimer's disease. They have also shown that PNMT protein is decreased in axon terminals in brains from patients with Alzheimer's disease; the decrease in PNMT appears to be due to retrograde degeneration of epinephrine neurons (Burke *et al.*, *Ann. Neurol.* 22:278-280 (1987)). In the case of advanced Alzheimer's disease, the Burke et al. presented evidence that the accumulation of PNMT in the perikarya results from diminished transport of this enzyme to axon terminals (Burke et al., *J. Am. Geriatr. Soc.* 38:1275-1282 (1990)).

Neurons that contain PNMT have cell bodies in brain stem regions of the rat brain and send projections mainly into other brain stem areas, such as the hypothalamus and the spinal cord. These neurons can be affected pharmacologically by various kinds of drugs. PNMT inhibitors currently represent the only means of modifying adrenaline neurons pharmacologically without affecting noradrenaline or dopamine neurons in brain. Experiments conducted in deoxycorticosterone

acetate-salt (DOCA-salt) hypertensive rats and spontaneously hypertensive rats (SHR) have shown that inhibitors of PNMT lower blood pressure (Goldstein *et al.*, Life Sci. 30:1951-1957 (1982); Lyang *et al.*, Res. Commun. Chem. Pathol. Pharmacol. 46:319-329 (1984); Chatelain *et al.*, J. Pharmacol. Exp. Ther. 252:117-125 (1990)). Molecules and compounds affecting adrenaline
5 neurons may also be of use in the treatment of psychiatric disorders and neuroendocrine dysfunction.

One embodiment of the subject invention provides polypeptides comprising the sequence of PNMT A. Other polypeptides of the invention include polypeptides comprising the amino acids of SEQ ID NO: 409 from positions 50 to 70, 83 to 103, 131 to 151 and/or 196 to 216. Also
10 encompassed by the instant invention are biologically active fragments of the PNMT A protein. "Biologically active fragments" are defined as those peptide or polypeptide fragments of PNMT A which have at least one of the biological functions of the PNMT A protein (e.g., the ability to catalyze the formation of adrenaline). In a preferred embodiment, the biologically active fragment of PNMT A contains at least one of the amino acid substitutions which distinguish PNMT A from
15 PNMT (*i.e.*, an isoleucine residue (I) substituted for a valine residue (V) at position 95; and/or valine residue (V) substituted for a glutamine residue (G) at position 130). In one embodiment, the PNMT A polypeptides of the invention are encoded by clone 179-1462-0-F11-CS.

Thus, one embodiment of the invention provides an enzymatic component of the adrenaline synthetic pathway and methods of producing adrenaline in accordance with methods known to those
20 skilled in the art. These methods substitute PNMT A, or biologically active fragments thereof, for the PNMT enzyme used in these known synthetic pathways.

The invention also provides variants of the protein of SEQ ID NO: 409. These variants have at least about 80%, more preferably at least about 90%, and most preferably at least about 95% amino acid sequence identity to the amino acid sequence of PNMT A. Variants according to the
25 subject invention also have at least one functional or structural characteristic of PNMT A, such as the ability to catalyze the formation of adrenaline. The invention also provides biologically active fragments of the variant proteins. Unless otherwise indicated, the methods disclosed herein may be practiced utilizing PNMT A or variants thereof. Likewise, the methods of the subject invention may be practiced using biologically fragments of PNMT A, or variants thereof, provides that said
30 biologically active fragments contain the amino acid substitutions noted supra.

One embodiment of the subject invention provides methods of using the protein of the invention, or biologically active fragments thereof, to label (chemically or isotopically) the adrenaline molecule *in vitro*. The labeled adrenaline molecules can then be used to localize receptors in tissue cuts by *in situ* hybridization experiments.

35 The invention also provides a fusion protein or polypeptide in which PNMT A, or biologically active fragments thereof, are combined with another protein (tag) by the use of a recombinant DNA molecule. The resulting purified, and enzymatically active fusion product, is then

added, *in vitro*, to the noradrenaline precursor and to S-adenosyl-L-methionine as a methyl donor. The enzymatic reaction is then performed in conditions known to those skilled in the art (Burke *et al.*, Proc. Soc. Exp. Biol. Med. 181:66-70 (1986); Morimoto *et al.*, Endocr. J. 40:179-183 (1993)). In this reaction, the methyl group of S-adenosyl-L-methionine must be labeled isotopically ([¹⁴C]-S-adenosyl-L-methionine or [methyl-³H]-S-adenosylmethionine), or chemically, in order to allow the transfer of a "tagged" methyl group to the adrenaline molecule.

Similarly, in cells transfected with cDNAs encoding the protein of the invention PNMT activity of expressed proteins may be measured by incubating cytosolic fractions with [¹⁴C]-S-adenosyl-L-methionine and normetanephrine for 60 min according to methods described by those skilled in the art (Morimoto *et al.*, *ibid.*). Agonists and/or antagonists of PNMT activity may also be tested (high throughput screening) on transfected cells expressing the wild type form of the protein of the invention. Again, effects of such drugs on PNMT enzymatic activity is measured by the methods described above.

The invention further relates to methods and compositions used to modify the protein of the invention (i.e. derivatize the PNMT A protein). Post-translational modifications encompassed by the invention include, N-linked or O-linked carbohydrate chains, processing of N-terminal or C-terminal ends, attachment of chemical moieties, such as polyethylene glycol, to the amino acid backbone, chemical modifications of N-linked or O-linked carbohydrate chains, and addition or deletion of an N-terminal methionine residue as a result of prokaryotic host cell expression. Some of these modifications of the protein of the invention may facilitate its extraction and purification in prokaryotic expression systems. Post-translational modifications such as N-linked or O-linked carbohydrate chains addition may also optimize the enzymatic activity of the protein of the invention when it is first produced in a prokaryotic system.

Another embodiment of the subject invention provides antibodies directed against the protein of the invention or immunogenic fragments thereof. The antibodies of the invention are useful for the screening of tissues and cells producing adrenaline or for affinity purification of PNMT or PNMT A. These antibodies may also be used in the diagnosis of pathologies and disorders such as pheochromocytomas and Alzheimer's disease, where PNMT A is overexpressed. Methods of performing affinity purification as well as methods of making polyclonal and monoclonal antibodies are well known to those skilled in the art.

In therapeutic regimens, neutralizing antibodies may be used as antagonists of PNMT A and used to treat conditions associated with overexpression of PNMT A. These disorders include, and are not limited to, hypertension, pheochromocytomas, and advanced Alzheimer's disease (Goldstein *et al.*, Life Sci. 30:1951-1957 (1982); Lyang *et al.*, Res. Commun. Chem. Pathol. Pharmacol. 46:319-329 (1984); Chatelain *et al.*, J. Pharmacol. Exp. Ther. 252:117-125 (1990); Isobe *et al.*, J. Urol. 163:357-362 (2000); Burke *et al.*, J. Am. Geriatr. Soc. 38:1275-1282 (1990)).

Proteins of SEQ ID NOs: 395 and 403 (internal designation: 160-101-3-0-H2-CS and 160-99-4-0-E4-CS respectively)

The 367-amino-acid-long proteins of SEQ ID NOs: 395 and 403 encoded by the cDNAs of SEQ ID NOs: 154 and 162 respectively are polymorphic variants, the first one being overexpressed
5 in fetal brain and ovary and the second one in fetal brain only. They both contain glutathione S-transferase (GST) domains from positions 47 to 122, and 206 to 309 which are respectively the G-site and H-site described below. In addition, they also display two hydrophobic domains (from aa 258 to aa 278 and from aa 338 to aa 358) which are characteristic of some GST proteins.

Glutathione S-transferase proteins (GSTs) are dimeric proteins that catalyse the conjugation
10 of glutathione to a wide range of hydrophobic compounds (through the formation of a thioether bond with their electrophilic centre) to create the products which are less reactive, more hydrophilic, and thus more easily excreted from the cells. The GST superfamily (E.C. 2.5.1.18) is indeed believed to be one of the most important proteins in the detoxification of reactive electrophiles within living cells. Glutathione is a cellular tripeptide (gamma-
15 glutamylcysteinylglycine) which is perhaps the most abundant amino acid derivative contained in the cells of higher life forms. The middle amino acid in glutathione, cysteine, has a free thiol group which can compete with the nucleophilic site on nucleotide bases for reaction with electrophiles. Within the cell, glutathione functions so as to conjugate to xenobiotic toxic molecules in general, and electrophiles in particular, to render the toxic molecules less reactive against cellular
20 macromolecules and to target the toxic molecules for subsequent metabolic and excretion pathways.

Based on amino acid sequence identity, there are at least seven major classes of GST proteins (designated alpha, kappa, mu, pi, sigma, theta and zeta). Sequence similarity between classes is rather low, ranging between 20-30%. However, a single point mutation in the H-subsite region of GST is enough to shift substrate specificity from class pi to alpha (Nuccetelli M.N. et al.
25 Biochem.Biophys.Res.Commun. 252: 184-189 (1998)). In spite of relatively low sequence identity, the GSTs exhibit a high degree of structural similarity. It is generally known that the GST molecule binds quite specifically and with high affinity to glutathione, but binds promiscuously to a wide variety of xenobiotic, electrophilic, and alkylating chemical agents. All GST enzymes of the four main cytosolic classes is found in dimeric form with two active sites per dimer each of which
30 functions independently of the other. The active site has been characterized as consisting of a glutathione binding region (designated the G-site) and a non-specific hydrophobic binding region (designated the H-site) to accommodate the electrophilic substances. Pi-, mu-, alpha- and theta-class crystal structures have been elucidated; all possess a similar GSH-binding site, but the hydrophobic substrate-binding site (H-subsite) is subject to variation across the classes (Allardyce C.S. et al.
35 Biochem.J. 343 525-531 (1999)). The GST activity has been suggested be involved in the regulation of the assembly of multisubunit complexes by shifting the balance between glutathione, disulfide glutathione, thiol groups of cysteines, and protein disulfide bonds. The GST domain is a

widespread, conserved enzymatic module that may be covalently or noncovalently complexed with other proteins. Regulation of protein assembly and folding may be one of the functions of GST (Koonin EV et al. Protein Sci 3:2045-2054 (1994)).

The cytosolic glutathione S-transferase are known to belong to four classes, designated
5 Alpha, Mu, Pi and Theta. A fifth class of glutathione S-transferases is a microsomal enzyme found primarily in liver endoplasmic reticulum. An extensive analysis of the expression microsomal glutathione transferase 1 in human tissues shows that it predominantly occurs in liver and pancreas. The relative expression levels in man ranged from: liver and pancreas to kidney, prostate, colon (30-40%), heart, brain, lung, testis, ovary, small intestine (10-20%), placenta, skeletal muscle, spleen,
10 thymus and peripheral blood leucocytes (1-10%). Liver-enriched expression was detected in human fetal tissues with lung and kidney displaying lower levels (10-20%). No transcripts could be detected in fetal brain or heart (Estonius M et al. Eur J Biochem 260:409-13 (1999)). Based on these observations, and the fact that the enzyme is encoded by a highly conserved single-copy gene, it is suggested that microsomal glutathione transferase 1 performs essential functions vital to most
15 mammalian cell types. One particular glutathione S-transferase was still identified in mitochondrial matrix (Pemble S.E. et al. Biochem.J. 319 : 749-754 (1996)).

GST and GST-like proteins are largely spread in organisms. In vertebrates and in cephalopodes some proteins (crystallins) presented in the lenses are structurally related to alpha-class GSTs (Chiou S.H. et al. Biochem.J. 309 : 793-800 (1995)). Furthermore, the olfactory
20 epithelial cytosol shows the highest GST activity among the extrahepatic tissues. The olfactory GSTs were found to catalyse glutathione conjugation of several odorant classes, including many unsaturated aldehydes and ketones, as well as epoxides and were proposed to play an important rôle in chemoreception (Ben-Arie N. et al. Biochem.J. 292 : 379-384 (1993)).

Higher cells each contain a family of many GST isozymes in each class with broad, yet
25 overlapping, specificity. Mu-class GSTs are thought to be involved in the detoxification of reactive oxygen species (cyclised o-quinones) produced via oxidative metabolism of catecholamines. These toxins are thought to be involved in neurological disorders of the nigrostriatal and mesolimbic systems (Parkinsons and Schizophrenia, respectively). Enzymes of the mu-class GSTs are expressed in the substantia nigra and have preferential substrate specificity for the cyclised o-quinones formed
30 by catecholamine metabolism (Hansson L.O. et al. J.Mol.Biol. 287: 265-276 (1999), Takahashi Y. et al. J. Biol. Chem. 268: 8893-8 (1993)). Whilst most of the GSTs share common substrates, there are distinct differences in substrate preference between subfamilies. These enzymes have evolved as a cellular protection system against a wide variety of electrophilic compounds, including a range of xenobiotics, oxidative metabolism by-products (oxidized lipid, DNA and catechols), and in
35 particular are known to metabolise a number of environmental carcinogens.

GSTs are also known to catalyze other reactions, such as peroxidase and isomerase reactions (Edwards R. et al. Trends in Plant Sci. 5 : 193-198 (2000)) as well as the addition of

aliphatic epoxides and arene oxides to glutathione; the reduction of polyol nitrate by glutathione to polyol and nitrite; certain isomerization reactions and disulfide interchange. As well, there are marked species differences in catalytic activities between various purified mammalian hepatic GST mixtures. Some of them catalyse chemical stereospecific conversion of several pharmacological substances much less effective than others. For example, recombinant human GST was successfully used in the reaction of steric conversion of 13-cis-retinoic acid to all-trans-retinoic acid (Chen H. and Juchau M.R. *Biochem.J.* 336 : 223-226 (1998)).

An increasing number of GST genes are being recognized as polymorphic. Certain alleles, particularly those that confer impaired catalytic activity may be associated with increased sensitivity to toxic compounds. Genetic polymorphisms and differences in GST expression have been implicated in individual susceptibility to certain types of cancer (for rev. Hayes JD and Strange RC *Pharmacology* 61:154-166 (2000)). For example, GSTM1 deficiency predisposes to head and neck cancer, especially to cancer of the larynx, which is particularly exposed to tobacco smoke carcinogens (Gronau S et al. *Laryngorhinootologie* 79:341-344 (2000)). Conversely, over-expression of GSTs is thought to be involved in the phenomenon of multi-drug resistance to cancer chemotherapy. One of the class of electrophilic compounds that are substrates for the glutathione S-transferase enzymes is the group of alkylating agents used in antineoplastic therapy. A common problem that is observed in modern cancer chemotherapy is the appearance of chemotherapeutic resistant tumor cells that, because of the resistivity, no longer respond appropriately to the antineoplastic agents. This resistance is often observed with many drugs that have no physical or mechanistic similarities to the original agent. GST isoenzymes have been shown to be involved in the development of drug resistance to a variety of chemotherapeutic agents such as adriamycin, vinblastine, actinomycin D and colchicine (Beckett, et al. *Adv. Clin. Chem.* 30:281-380 (1993)). It has been demonstrated that a resistant population of malignant cells shows a modified pattern of total glutathione S-transferase activity. A resistant population of MCF-7 breast cancer cells, identified through selection in adriamycin by Batist et al., *J. Biol. Chem.*, 261:15544-15549 (1986) resulted in a subset of cells which were approximately 200 fold more resistant than the parental cells. The resistant cells were found to exhibit a 45 fold increase in total glutathione S-transferase activity, the increase being due to the result of an appearance of an isozyme not expressed in the parental cell line. It was demonstrated that an increase in glutathione S-transferase alone, an increase conditioned by the transformation of susceptible cells with a foreign DNA construct expressing the wild-type glutathione S-transferase coding region, could increase the resistance of cells to an antineoplastic agent. As reported in Puchalski and Fahl, *Proc. Natl. Acad. Sci. USA*, 87:2443-2447 (1990), expression of the rat 1-1, 3-3 and the human P1-1 isozymes of glutathione S-transferase in COS cells increased their resistance to the agent. The recent study of increase in resistance of tumor cells to cytotoxic drugs or ionizing radiation has allowed to identify using differential display a new GST-related protein p28 expressed exclusively in lymphoma cell (Kodym

R. et al. *J. Biol.Chem.* 274: 511-5137 (1999)). Subcellular protein fractionation revealed p28 localization in the cytoplasm, but with thermal stress p28 relocated to the nuclear fraction of cellular proteins. The sequence homology and the similar functional characteristics of p28 to other GST family members (in particular relocation in response to thermal stress and ability to bind glutathione), argues that p28 is a new mammalian member of GST superfamily.

Evidence suggests that the level of expression of GSTs is a crucial factor in determining the sensitivity of cells to a broad spectrum of toxic chemicals. In humans, marked interindividual differences exist in the expression of class alpha, mu and theta GST. For the most abundant mammalian classes of GST the mechanisms of transcriptional and post-translational regulation have been studied. The biological control of alpha-, mu- and pi- classes exhibit sex-, age-, tissue-, species-, and tumor-specific patterns of expression. In addition, GST are regulated by a structurally diverse range of xenobiotics and, to date, more than 100 chemicals have been identified that induce GST (Hayes J.D. and Pulford D.J. *Crit.Rev.Biochem.Mol.Biol.* 30 : 445-600 (1995)). A significant number of these chemicals occur naturally and, as they are found as nonnutrient components in vegetables and citrus fruits, it is apparent that humans are likely to be exposed regularly to such compounds. Many inducers effect transcriptional activation of GST genes through either the antioxidant-responsive element (ARE), the xenobiotic element (XRE), the GST P enhancer 1 (GPE), or the glucocorticoid-responsive element (GRE). Many of compounds that induce GST are themselves substrates for these enzymes, or are metabolized (by cytochrome P-450 monooxygenases) to compounds that can serve as GST substrates, suggesting that GST induction represent part of an adaptive response mechanism to chemical stress by electrophiles. It also appear probable that GST are regulated in vivo by reactive oxygen species, the potents inducers capable of generating free radicals by redox-cycling ; such relulation can be an adaptive response to oxydative stress in the cell. It has been shown GST-pi can potently and selectively inhibit activation of jun protein by its upstream kinase (JNK) ; these results suggest GST-pi can also be a regulator of signal transduction (Monaco R. et al. *J.Prot.Chem.* 18 : 859-866 (1999)). The majority of human tumors express significant amounts of class pi GST (Hayes&Pulford, supra).

Therefore, GSTs have medical importance due to their role in mediating drug resistance in cancer patients. The measurement of GST isoenzymes in vitro has importance in diagnostic medicine. For example, the measurement of the pi isoenzyme of GST in tissue specimens is useful in pathology for the detection and diagnosis of a variety of different tumors. In addition, measurement of the alpha form of GST in blood is useful for the detection and monitoring of a variety of different forms of liver disease (for a detailed description of the clinical applications of GST measurements see Beckett, et al., supra).

It is believed that the proteins of SEQ ID NOs: 395 and 403 or part thereof are transferases, probably transferring alkyl or acyl groups different from methyl group, more probably glutathione S-transferases and, as such, play a role in cellular detoxification especially against xenobiotics and

oxidative metabolism byproducts. Preferred polypeptides of the invention are polypeptides comprising the amino acids of SEQ ID NOs: 395 and 403 from positions 47 to 122, and 260 to 309. Other preferred polypeptides of the invention are fragments of SEQ ID NOs: 395 and 403 having any of the biological activities described herein. The transferase activity of the proteins of the invention or part thereof may be assayed using any of the assays known to those skilled in the art including those described for GST proteins as in US patents 5,866,792 and 6,096,504, which disclosures are hereby incorporated by reference in their entireties.

To find substrates, the proteins of the invention, or part thereof, or derivative thereof, may be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes, between the proteins of the invention, or part thereof, or derivative thereof, and the agent being tested, may be measured. Antagonists or inhibitors of the proteins of the invention may be produced using methods which are generally known in the art, including the screening of libraries of pharmaceutical agents to identify those which specifically bind the protein of the invention. Another technique for drug screening which may be used provides for high throughput screening of compounds having suitable binding affinity to the proteins of the invention as described in published PCT application WO84/03564.

The invention relates to methods and compositions using the proteins of the invention or part thereof or derivative thereof to catalyze GST-dependent detoxification reactions in vitro or in vivo using any methods known to those skilled in the art. For example, uses of the proteins of the invention or part thereof may be very useful to treat toxic byproducts such as the ones obtained in laboratory experiments, such as dietary toxins due to the use of pesticides on plants used to feed animals or humans, etc... Preferably, the proteins of the invention or part thereof or derivative thereof is added to a sample containing the substrate(s) in conditions allowing detoxification, and allowed to catalyze the detoxification of the substrate(s). In a preferred embodiment, the detoxification is carried out using a standard assay such as those described herein.

In some of the above cited embodiments, compositions comprising the proteins of the present invention or part thereof are added to samples as a "cocktail" with other detoxifying enzymes. The advantage of using a cocktail of detoxifying enzymes is that one is able to detoxify a wide range of substrates without knowing the specificity of any of the enzymes. Using a cocktail of detoxifying enzymes also protects a sample from a wide range of future unknown toxic compounds from a vast number of sources. For example, the proteins of the invention or part thereof is added to samples where toxic compounds are undesirable. Alternatively, the protein of the invention or part thereof may be bound to a chromatographic support, either alone or in combination with other detoxifying enzymes, using techniques well known in the art, to form an affinity chromatography column. A sample containing the undesirable substrate is run through the column to remove the substrate. Immobilizing the proteins of the invention or part thereof on a support is particularly

advantageous for those embodiments in which the method is to be practiced on a commercial scale. This immobilization facilitates the removal of the enzyme from the batch of product and subsequent reuse of the enzyme. Immobilization of the protein of the invention or part thereof can be accomplished, for example, by inserting a cellulose-binding domain in the protein. One of skill in the art will understand that other methods of immobilization could also be used and are described in the available literature. Alternatively, the same methods may be used to identify new substrates.

In a preferred embodiment, the invention relates to cells and plants or animals genetically engineered to express the protein of the invention or part thereof, preferably at a high level using any method known to those skilled in the art. Such engineered cells, animals or plants will display enhanced detoxification of compounds. In a more preferred embodiment, expression of the proteins of the invention or part thereof will confer resistance to herbicides to transgenic plants using techniques similar to those described in the US patent 5,866,792.

For such embodiments, the proteins of the invention may need to be modified to enhance their ability to react with specific substrates. These modifications can provide novel isoforms which are specifically efficient against selected electrophilic or alkylating agents. Artificial DNA constructs encoding and expressing such modified or mutant proteins of the invention or part thereof may be selectively delivered into targeted cells to enhance the resistivity of those cells to the alkylating or neoplastic agents. The methods related to such modifications are described (Fahl, et al. United States Patent 6,136,605 Oct24, 2000). The method is based on random mutation and selection with the selection being performed with the agent against which enhanced activity is sought. The mutation is preferably site directed to the amino acids associated with the H-site on the enzyme, so as to favor the creation of new, useful isoforms of the enzyme.

In another embodiment, the invention relates to compositions and methods using the proteins of the invention or part thereof to design specific systems of artificial chemoreception as described in Ben-Arie N. et al, supra. Such chemoreception systems could recognize odorants, xenobiotics, pesticides, drugs and may be useful for chemical, cosmetic, pharmaceutical, forensic and any other analytical purposes. The design of such a system may be generally based on the subtle specificity of recognition of compounds by different GST isoenzymes. The methods to produce analytical diagnostics based on enzyme specificity are known by those skilled in the art.

In another embodiment, the invention relates to compositions and methods using the proteins of the invention or part thereof such as ligands for substrates of interest. In a preferred embodiment, the proteins of the invention or part thereof may be used to identify and/or quantify substrates using any techniques known to those skilled in the art such as those described in Koonin E., supra. In another preferred embodiment, the proteins of the invention or part thereof may be used to improve or to modify some molecular biology methods based on protein-protein interactions, including but not limited to two-hybrid assays, expression and purification systems

based on GST fusion to heterologous proteins as already available commercially (expression vectors or plasmids encoding fusions proteins and affinity purification methods).

In still another embodiment, the invention relates to methods and compositions using the proteins of the invention or part thereof as a marker protein to selectively identify tissues, preferably fetal brain for the protein of SEQ ID NO: 403, and preferably fetal brain and ovary for the protein of SEQ ID NO: 395. For example, the protein of the invention or part may be used to synthesize specific antibodies using any techniques known to those skilled in the art including those described therein. Such tissue-specific antibodies may then be used to identify tissues of unknown origin, for example, forensic samples, differentiated tumor tissue that has metastasized to foreign bodily sites, or to differentiate different tissue types in a tissue cross-section using immunochemistry.

In another embodiment of the invention, measurement of the activity or expression of the proteins of the invention, may be used for the assessment of organ status, including organ damage following immunological or toxological insult and diagnostic of transplant rejection, using any technique known to those skilled in the art including those described in US patents 6,080,551 and RE35,419.

In another embodiment of the invention, the proteins of the invention or part thereof, or derivative thereof, may be used to diagnose, treat and/or prevent cell proliferative disorders linked to dysregulation of gene expression of the proteins of the invention. Such disorders include but are not limited to, benign tumors, and cancers such as adenocarcinoma; leukemia; melanoma; lymphoma; sarcoma; and cancers of the brain, ovary, bladder, colon, liver, small intestine, large intestine, breast, kidney, lung, and prostate. Diagnosis may be performed using nucleic acids or antibodies able to detect the expression of the protein of the invention using any technique known to those skilled in the art including Northern blotting, RT-PCR, immunoblotting methods immunohistochemistry, enzyme-linked immunosorbant assay (ELISA) described herein. Quantities of the protein of the invention expressed in subject samples, control and disease from biopsied tissues or body fluids or cell extracts taken from patients are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

For prevention and/or treatment purposes, the expression of the proteins of the invention may be enhanced using any methods known to those skilled in the art. For example, gene therapy techniques may be used such as the delivery of sense promoter polynucleotide constructs for the proteins of the invention or part thereof using a recombinant expression vector such as a chimeric virus or a colloidal dispersion system (see Nelson et al. United States Patent 552,277).

Alternatively, the proteins of the invention or fragments thereof or derivatives thereof may be administered to a subject to treat or prevent cancerous and precancerous disorders as well as a proliferative disorders in general. Such disorders can include, but are not limited to, syndromes represented by abnormal neoplastic, including dysplastic, changes of tissue, dysplastic growths in ovary, brain, colonic, breast, prostate or lung tissues, dysplastic nevus syndromes, polyposis

syndromes, colonic polyps, precancerous lesions of the cervix (i.e., cervical dysplasia), esophagus, lung, prostatic dysplasia, prostatic intraneoplasia, breast and/or skin and related conditions (e.g., actinic keratosis), whether the lesions are clinically identifiable or not.

The invention also relates to compositions and methods using the proteins of the invention or part thereof or derivative thereof to decrease drug resistance in cancer chemotherapy. Inhibition of the expression and/or activity of the proteins of the invention may be achieved using any mean known to those skilled in the art. In a preferred embodiment, gene therapy methods such as antisense oligonucleotides, triple helices strategies are described elsewhere in the application. In another preferred embodiment, antagonists of the activity of the proteins of the invention may be used. These antagonists may be directly administered to patients. Low-molecular-weight inhibitors (i.e., those which can be delivered freely into the brain and which specifically inhibit GST activity) are especially preferred for use in cancer therapy. Alternatively, artificial DNA constructs encoding peptide modulators or inhibitors of the activity of the protein of the invention and flanking sequences effective to express the protein coding sequence in a host cell as well as flanking regulatory sequences (such as an antioxidant responsive element which enhances the expression of the glutathione S-transferase in the presence of antioxidant molecules) may be used. Such artificial DNA constructs confers to recombinant cells an increased level of resistance to an antineoplastic agent.

There is a need for selective inhibitors of GST isoenzymes for treatment of drug resistance in cancer patients. Thus, in a further embodiment of this invention, the proteins of the invention or fragments thereof can be used for screening of the compounds which are selective inhibitors or specific inhibitors of one or more GST isoenzymes. Selective inhibition means that a compound has a greater inhibitory effect on one isoenzyme than it does on another GST isoenzyme. Such compounds could also be tested and selected for their ability to overcome drug resistance to chemotherapeutic agents (see Jones, et al. United States Patent 6,103,665 Aug 2000). For example, mammalian cell lines that have been made resistant to particular chemotherapeutic drugs can be used to identify haloenol lactone compounds that render the lines sensitive to the chemotherapeutic agents. Such cell lines are known to those of skill in the art and can be obtained for example from the American Type Culture Collection, Rockville, Md., USA.

Protein of Seq Id No: GRP (187-38-0-0-110-CS)

The protein of SEQ ID No: 306, herein referred as GRP, encoded by the cDNA of SEQ ID No: 65 herein referred as GRP2, is homologous to bovine glutamic-acid rich protein (GARP) (GENEPEPT ID: M61185). The protein of the invention is overexpressed in the brain and fetal brain, lymph ganglia and thyroid.

The protein of the invention exhibits homology with bovine glutamic-acid rich protein (GARP) (18 % identical amino acids, 28% positive amino acids when aligned by BLASTP 2.0.9).

GARP proteins have been identified as multivalent proteins that interact the key players of cGMP signaling, phosphodiesterase and guanylate cyclase. GARP proteins are closely associated to cyclic nucleotide-gated channels (CNGs) which make up a family of nonselective cation channels found in a variety of tissues. The beta subunit of CNGs have a unique bipartite structure, containing a membrane-spanning region (beta part) and a GARP part (GARP). GARP is highly homologous to a soluble splice form, GARP1, and a splice variant lacking the C-terminal glutamic-acid-rich region. Experiments using GARP attached to affinity columns showed that phosphodiesterases are highly retained by the column [Korschen HG et al., Nature, 400(6746):761-766 (1999)]. Moreover, Korschen et al. demonstrated that GARP inhibits both soluble and membrane bound phosphodiesterase.

Cyclic nucleotides are involved in regulating the activity of airway smooth muscle and many other cells in the airways, including pro-inflammatory, immunocompetent cells such as macrophages, eosinophils, mast cells and lymphocytes. Cyclic nucleotides are inactivated by the action of cyclic nucleotide phosphodiesterase enzymes (PDE). Inhibition of cGMP PDE results in elevation of cGMP levels; elevated cGMP levels are associated with beneficial anti-platelet, anti-neutrophil, anti-vasospastic and vasodilatory activity.

Thus, the subject invention provides a polypeptide having the sequence of SEQ ID No: 306 or a GRP polypeptide encoded by the human cDNA of clone 187-38-0-0-110. In a preferred embodiment, GRP is encoded by the sequence of SEQ ID No: 65 or the human cDNA of clone 187-38-0-0-110, however, all polynucleotides encoding the polypeptides of the invention are included. As used herein, "the GRP protein" includes the full length protein of SEQ ID NO: 306 as well as biologically active fragments of the GRP protein. Also encompassed by the phrase "the GRP protein" are variants of the protein of SEQ ID NO: 306 and biologically active fragments of said variant proteins.

"Biologically active fragments" are defined as those peptide or polypeptide fragments of GRP which have at least one of the biological functions of the full length protein (e.g., the ability to inhibit the activity of PDE or serve as an affinity substrate for PDEs).

The invention also provides variants of the protein of the GRP protein encoded by SEQ ID NO: 306. These variants have at least about 80%, more preferably at least about 90%, and most preferably at least about 95% amino acid sequence identity to the amino acid sequence of GRP. Variants according to the subject invention also have at least one functional or structural characteristic of GRP, such as the ability to inhibit the activity of PDE or serve as an affinity substrate for PDEs. The invention also provides biologically active fragments of the variant proteins. Unless otherwise indicated, the methods disclosed herein can be practiced utilizing GRP or variants thereof. Likewise, the methods of the subject invention can be practiced using biologically active fragments of GRP, or biologically active fragments of GRP variants.

Assays related to the inhibitory effect of the protein of the invention can be carried out using techniques described in U.S. Patent No. 6,130,333, hereby incorporated by reference in its entirety; or by any other technique known to those skilled in the art.

One aspect of the subject invention provides compositions and methods of using the
5 nucleotide sequence of SEQ ID NO: 65, or its complement, in molecular biology techniques. These techniques include, but are not limited to: the use of segments of GRP2 as oligomers for PCR; expression of the GRP2 and the production of recombinant proteins; in generation of antisense RNA and DNA, their chemical analogs and the like; the use of GRP2 segments as hybridization probes and in chromosome gene mapping.

10 For example, nucleotide sequence of SEQ ID No: 65, or its complement, can be used to generate hybridization probes for mapping the naturally occurring genomic sequence. The sequence can be mapped to a particular chromosome or to a specific region of the chromosome using well known techniques. These include in situ hybridization to chromosomal spreads, flow-sorted chromosomal preparations, or artificial chromosome constructions such as yeast artificial
15 chromosomes, bacterial artificial chromosomes, bacterial P1 constructions, or single chromosome cDNA libraries as reviewed in Price (Price CM – Blood Rev. – 1993, 7(2):127-34) and Trask B (Trask BJ – Trends Genet. – 1991, 7(5):149-54).

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, are invaluable in extending genetic
20 maps; genetic maps provide valuable information to investigators searching for disease-causing genes using positional cloning or other gene discovery techniques. The nucleotide sequence of the present invention can also be used to detect differences in the chromosomal location due to translocation, inversion, etc. among normal, carrier or affected individuals.

Another embodiment of the subject invention provides pharmaceutical compositions
25 comprising the GRP protein and pharmaceutically acceptable carriers. These pharmaceutical compositions can be used in prophylaxis and/or treatment of a variety of conditions where inhibition of phosphodiesterase is considered to be beneficial. The biochemical, physiological, and clinical effects of phosphodiesterases inhibitors suggest their utility in a variety of disease states in which modulation of smooth muscle, renal, hemostatic, inflammatory, and/or endocrine function is
30 desirable. Therefore, the GRP protein can be used for the treatment or prophylaxis of a number of disorders and conditions including, but not limited to, stable, unstable, and variant (Prinzmetal) angina; hypertension; pulmonary hypertension; congestive heart failure; acute respiratory distress syndrome; acute and chronic renal failure; atherosclerosis; conditions of reduced blood vessel patency (e.g., postpercutaneous transluminal coronary or carotid angioplasty, or post-bypass surgery
35 graft stenosis); peripheral vascular disease; vascular disorders, such as Raynaud's disease, thrombocythemia, intermittent claudication; immune diseases, multiple sclerosis; cancers inflammatory diseases, graft versus host disease, Alzheimer's disease, memory deficits, , stroke,

bronchitis, chronic asthma, acute lung injury, chronic obstructive pulmonary disease, allergic asthma, allergic rhinitis; glaucoma; osteoporosis; preterm labor; benign prostatic hypertrophy; male and female erectile dysfunction; and diseases characterized by disorders of gut motility (e.g., irritable bowel syndrome).

- 5 The GRP protein of the invention can also provide beneficial anti-platelet, anti-neutrophil, anti-vasospastic, vasodilatory, natriuretic, and diuretic activities when administered in therapeutically effective amounts. The GRP protein can also potentiate the effects of endothelium-derived relaxing factor (EDRF), gastric NO administration, nitrovasodilators, atrial natriuretic factor (ANF), brain natriuretic peptide (BNP), C-type natriuretic peptide (CNP), and endothelium-
- 10 dependent relaxing agents such as bradykinin and acetylcholine, when administered to an individual.

Another embodiment of the subject provides methods of treating male erectile dysfunction comprising the administration of therapeutically effective amounts of the GRP protein using appropriate methods known to the skilled artisan.

- 15 Another embodiment of the subject invention provides industrially significant methods of recovering PDE comprising contacting solutions containing PDE with immobilized GRP protein. In this aspect of the invention, the GRP protein is immobilized onto a solid support and allowed to specifically bind to PDE contained in a solution or sample. PDE can then be eluted from the immobilized GRP protein according to methods known to the skilled artisan (see, for example,
- 20 Korschén et al., supra). PDE is a commercially valuable commodity sold by various vendors.

Protein of SEQ ID NO: 302 (internal designation 187-2-2-0-A3-CS)

The protein of SEQ ID NO: 302 encoded by the cDNA of SEQ ID No: 61 is related to a neuronally expressed protein (neuritin, Genseq accession number W37859) known to have a role in neuritogenesis and axonal and dendritic growth.

- 25 The 164 amino acid protein of SEQ ID NO: 302 is 24% identical to neuritin over the complete sequence. Specifically, SEQ ID NO: 302 displays two blocks of strong homology to neuritin (amino acids 41-60 of SEQ ID NO: 302 display 55% identity and 95% similarity to amino acids 30-49 of neuritin, and amino acids 66-117 of SEQ ID NO: 302 display 32% identity and 57% similarity to amino acids 62-113 of neuritin). The C-terminal portion of neuritin (aa 116-142 of
- 30 neuritin) is highly hydrophobic and contains a cleavage site found in GPI-anchored proteins. The protein of SEQ ID NO: 302 also has a hydrophobic C-terminus (21 out of the last 30 amino acids are hydrophobic) and conforms to the GPI anchor consensus sequence.

- Neuritin, also known as candidate plasticity-related gene number 15 (cpg-15), was independently identified by two groups from differential cDNA libraries generated from kainic
- 35 acid-treated hippocampal cells (Nedivi et al., Nature. 363:718-22 (1993); Naeve et al., Proc. Natl. Acad. Sci. USA. 94:2648-53 (1997)). Neuritin is a secreted protein that contains a potential GPI

anchoring domain believed to anchor the protein to the membranes of target cells. Neuritin is expressed strongly in the brain, and in particular, in systems with pronounced developmental plasticity, including the pyramidal neurons of the cornu ammonis and the granule cells of the hippocampus dentate gyrus. Strong expression is also observed in layers of tenia tecta projecting to the olfactory bulb, the major target of the retinal ganglion cells, and the optical nerve layer of the superior colliculus (optic tectum); and localized expression is observed in the thalamic nuclei and the cerebral cortex (Nedivi et al., *Proc. Natl. Acad. Sci. USA.* 93:2048-53 (1996); Naeve et al., *Proc. Natl. Acad. Sci. USA.* 94:2648-53 (1997); Nedivi et al., *Science.* 281:1863-66 (1998)). mRNA is expressed throughout development and persists into adulthood. In addition, neuritin expression is upregulated in adults by brain derived neurotrophic factor (BDNF). Neuritin mRNA is also detected in the lung and the liver, although at lower levels than that observed in the CNS (Naeve et al., *Proc. Natl. Acad. Sci. USA.* 94:2648-53 (1997)).

Functional studies on the neuritin protein have revealed a role in neuronal growth. In one such study, rat cortical and hippocampal neurons were treated with recombinant forms of neuritin. Neurons treated with neuritin showed extensive neuritogenesis over control cultures. Specifically, neurons showed well-differentiated cell bodies with well-defined extensions after treatment with neuritin (Naeve et al., *Proc. Natl. Acad. Sci. USA.* 94:2648-53 (1997)). Other studies using frog optic tectum showed that transfection of tectum cells with neuritin cDNA can increase the growth rate of tectal cell dendrites (Nedivi et al., *Science.* 281:1863-66 (1998)). Studies have also shown that neuritin can modify the growth of retinotectal axons by increasing the elaboration of presynaptic axons and can promote the maturation of retinal tectal synapses (Cantalupo et al., *Nature Neuroscience.* 3:1004-1011 (2000)). Together, these results indicate that neuritin promotes the growth of pre- and post-synaptic neurons and contributes to the formation and stabilization of mature synapses.

The subject invention provides the polypeptide of SEQ ID NO: 302 and polynucleotide sequences encoding the amino acid sequence of SEQ ID NO: 302. In one embodiment, the polypeptides of SEQ ID NO: 302, including fragments, variants, etc. are replaced by the corresponding polypeptide encoded by the human cDNA of clone 187-2-2-0-A3-CS. Also included in the invention are biologically active fragments of the protein of SEQ ID NO: 302 and polynucleotide sequences encoding these biologically active fragments. In another embodiment, biologically active fragments comprise amino acid positions 41-60, 66-117, 41-117, and 41-164. In another embodiment, these fragments may be joined together by chemical linkers or by recombinantly inserted amino acid linker segments according to methods known in the art. "Biologically active fragments" are defined as those peptide or polypeptide fragments of SEQ ID NO: 302 which have at least one of the biological functions of the full length protein (e.g., the ability to stimulate neuritogenesis and axonal and dendritic growth).

The invention also provides variants of SEQ ID NO: 302. These variants have at least about 80%, more preferably at least about 90%, and most preferably at least about 95% amino acid sequence identity to the amino acid sequence of SEQ ID NO: 302. Variants according to the subject invention also have at least one functional or structural characteristic of SEQ ID NO: 302, such as the biological functions described above. The invention also provides biologically active fragments of the variant proteins. Unless otherwise indicated, the methods disclosed herein can be practiced utilizing the polypeptide of SEQ ID NO: 302 or variants thereof. Likewise, the methods of the subject invention can be practiced using biological fragments of the protein of SEQ ID NO: 302 or variants of said biologically active fragments.

Because of the redundancy of the genetic code, a variety of different DNA sequences can encode SEQ ID NO: 302. It is well within the skill of a person trained in the art to create these alternative DNA sequences, which encode proteins having the same, or essentially the same, amino acid sequence. These variant DNA sequences are, thus, within the scope of the subject invention. As used herein, reference to "essentially the same sequence" refers to sequences that have amino acid substitutions, deletions, additions, or insertions that do not materially affect biological activity. Fragments retaining one or more characteristic biological activity of SEQ ID NO: 302 are also included in this definition.

"Recombinant nucleotide variants" are alternate polynucleotides which encode a particular protein. They can be synthesized, for example, by making use of the "redundancy" in the genetic code. Various codon substitutions, such as the silent changes which produce specific restriction sites or codon usage-specific mutations, can be introduced to optimize cloning into a plasmid or viral vector or expression in a particular prokaryotic or eukaryotic host system, respectively.

The protein of SEQ ID NO: 302, and variants thereof, can be used to produce antibodies according to methods well known in the art. The antibodies can be monoclonal or polyclonal. Antibodies can also be synthesized against immunogenic fragments of SEQ ID NO: 302, as well as variants thereof, according to known methods. The subject invention also provides antibodies which specifically bind to biologically active fragments of SEQ ID NO: 302 or biologically active fragments of SEQ ID NO: 302 variants.

The protein of SEQ ID NO: 302 can be utilized to treat diseases and disorders of the central or peripheral nervous system which arise from alterations in the pattern of expression of the protein of SEQ ID NO: 302. In this aspect of the subject invention, compositions comprising the protein of SEQ ID NO: 302 and a pharmaceutical carrier are administered to an individual in need thereof. Alternatively, in cases where the protein of SEQ ID NO: 302 is overexpressed, reductions in SEQ ID NO: 302 levels may be accomplished by a variety of methods known to those of skill in the art. These methods include the introduction of neutralizing antibodies or the use of antisense polynucleotides derived from the protein of SEQ ID NO: 302 or clone 187-2-2-0-A3-CS.

The subject invention also provides materials and methods for the treatment of neurological disorders comprising contacting neuronal cells with compositions comprising the protein of SEQ ID NO: 302 and pharmaceutically acceptable carriers. Thus, this aspect of the invention provides methods of treating patients suffering from a variety of neurological disorders, conditions, and/or diseases of the central, autonomic, or peripheral nervous system. These include neurological damage arising from congenital disease, trauma, surgery, stroke, ischemia, infection, metabolic disease, nutritional deficiency, malignancy, and/or exposure to toxic agents. Additional examples of such disorders include, but are not limited to, epilepsy, cerebral neutralisms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases, Creutzfeldt-Jakob disease, Gerstmann-Sträussler-Scheinker syndrome, fatal familial insomnia, diabetes induced peripheral neuropathy or neuropathy induced by other metabolic disorders or nutritional deficiencies, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder, akathisia, amnesia, and/or other dystrophies or degenerative disorders of the visual, sensory, olfactory, auditory, motor, or memory systems. Methods of introducing therapeutic compounds into cells are known to those skilled in the art. Non-limiting examples include the use of targeted liposomes, fusogenic liposomes, or other carriers suitable for the introduction of a therapeutic compound into a target cell.

The subject invention also provides materials and methods for the treatment of neurological disorders comprising contacting neuronal cells with compositions comprising polynucleotides encoding the protein of SEQ ID NO: 302 and pharmaceutically acceptable carriers. In one embodiment, the polynucleotide is clone 187-2-2-0-A3-CS. Methods of introducing polynucleotides into cells and directing expression of the polynucleotide are known to those skilled in the art.

Antibodies raised against the protein of SEQ ID NO: 302 may be used in a variety of immunoassays known to those skilled in the art. In this aspect of the invention, immunoassay screening for abnormal levels of the protein of SEQ ID NO: 302 can be used as screens or diagnostic/prognostic indicators of neurodegenerative disease.

Antibodies raised against the protein of SEQ ID NO: 302, fragments, and/or derivatives thereof may also be used for detection and identification of growing and differentiating neurons including, but not limited to, the pyramidal neurons of the cornu ammonis, the granule cells of the hippocampus dentate gyrus, neurons in layers of tenia tecta projecting to the olfactory bulb, the optical nerve layers of the superior colliculus (optic tectum), and neurons of the thalamic nuclei and the cerebral cortex.

Protein of SEQ ID NO:301 (187-12-4-0-A8-CS)

The protein of SEQ ID No:301, encoded by the cDNA of SEQ ID NO:60, is homologous to the Eukaryotic cell growth inhibiting factor (GENESEQP: R95950) described in patent WO9617933. The protein of the invention is highly expressed in the brain, fetal brain, fetal liver and the prostate.

It is believed that the protein of the invention is a cell growth inhibiting factor. Preferred polypeptides of the invention are those that comprise amino acids 221 to 287. Other preferred polypeptides of the invention are any fragment of SEQ ID NO:301 having any of the biological activities described herein. In the present invention, a cell inhibiting factor is defined as a peptide or protein that decreases, suppresses or terminates (reversibly or irreversibly) the growth of at least one type of cell such as, but not limited to, bacteria, yeast, vertebrate cells, mammalian cells and human cells, under ordinary culturing conditions known to those skilled in the art. Assay of the inhibiting activity of the invention can be carried out, for example, by evaluating the decrease in DNA synthesis as described in Patent WO 96/17933, or by measuring the number or density of cells using any standard method. For example, fibroblasts are transfected with a vector containing the DNA sequence coding for the protein of the invention or part thereof. Cells are then cultured in a standard medium, exposed to tritiated thymidine, and further cultured. The cultures are then fixed and stained with X-Gal, the blue stained galactosidase-expressing cells are counted under a microscope, and the ratio of cells showing dark particles in their nuclei due to tritiated thymidine uptake is determined. DNA synthesis inhibitory rates are calculated with the labeling index taking for reference (i.e. 0% inhibition) a culture of cells transfected with a "blank" vector (i.e. not modified to contain the DNA coding for the protein of the invention or part thereof).

Aging at the cell level is associated with individual aging. The maximum possible number of divisions (division life span) of cultured cells is inversely proportional to individual age. Even if an aged cell is fused with a young or immortalized cell, DNA synthesis does not occur again in aged cells; on the contrary, DNA synthesis in the young and immortalized cell is suppressed (Stein GH et al. – Proc Nat Acad Sci. – 1981, 78:p3025). This demonstrates that certain factors controlling cellular senescence are dominant, and that aged cells not only lack substances essential for their growth but also have substances that actively suppress DNA synthesis. Moreover, microinjections of mRNA, prepared from an aged cell, are known to inhibit DNA synthesis (Lumpkin CK et al. –

Science – 1986, 232:p393). Therefore as cells age, there are some genes that are newly expressed or whose expression is increased. Such genes play an important role, directly or indirectly, in cell aging.

Pereira-Smith et al. tested the complementation of a large number of immortalized human
5 cells in fused pairs and demonstrated the presence of 4 groups of human aging genes (Pereira-Smith
OM et al. – Proc Nat Acad Sci. – 1988, 85:p6042). Clarifying the nature of aging-associated genes
is not only important in understanding aging, both at the cellular and individual levels, but is also
significant in that the use of these genes or gene products would enable the diagnosis of various
aging-associated diseases and diseases caused by cellular senescence, the development of
10 prophylactic/therapeutic drugs for such diseases, and their application as prophylactic/therapeutic
drugs for various diseases involving uncontrolled cell growth such as, but not limited to, cancers.

In one embodiment of the present invention, the polypeptides and polynucleotides of the
invention are used to specifically label cells of the brain, fetal brain, fetal liver and the prostate, as
the protein is strongly expressed in these tissues. The ability to specifically detect these tissues, and
15 cells derived from these tissues, has a number of uses, including for the determination of the history
of tumor cells and for histological analyses.

An embodiment of the present invention relates to methods and compositions of
using the protein of SEQ ID NO:301 or the cDNA of SEQ ID NO:60 or any part thereof, to inhibit
cell proliferation in vitro. For example, by including the invention in a “cocktail” with other
20 proteins (such as proteases) it could be used as a decontaminant, i.e. to prevent the growth of any
cells to maintain a sterile environment. Preferred applications of this embodiment include
decontamination of samples (such as cell culture media) and instruments (such as surgical
instruments), where the invention would be used as a bacteriostatic/mycostatic agent. Another
example pertains to the use of the protein of the invention as a reagent for terminating the cell cycle
25 of cultured cells at a given time point, e.g., as a reagent for synchronizing cell division, avoiding the
need to isolate specific cells (e.g. at the desired cell cycle phase) in cultures using techniques such
as flow cytometry. Synchronization of the cell cycle could, for example, be achieved, e.g., by
transfecting cells with an appropriate vector containing the DNA coding for the protein of the
invention or part thereof, where expression of the protein results in growth inhibition. Then, after a
30 certain time, an inhibitor of the protein could be administered in order to enable the cells to resume
growth (e.g. all at the S phase, when DNA is synthesized). Further, the ability to synchronize the
cell cycle in an in vitro experimental system would provide improved assay precision or would
facilitate any laboratory procedure or experiment involving a particular cell cycle stage. Use of the
invention for in vitro inhibition of cell proliferation is not limited to the above examples; the
35 invention is potentially useful in any in vitro application that requires the inhibition of cellular
proliferation.

Another embodiment of the invention pertains to the introduction of SEQ ID NO:301 or SEQ ID NO:60, or any part thereof, into a target tissue (such as skin or vascular endothelium) to establish an in vitro aged cell line of the target tissue. Such a cell line is useful as a screening system for clarifying the mechanisms of aging and/or cellular senescence, but also for seeking prophylactic and therapeutic drugs for aging-associated diseases and or diseases caused by cellular senescence, as it provides an in vitro model of aged cells (in vitro cell cultures provide the advantage of being easily produced and are not as expensive as animal models). Thus these cells are potentially useful in drug candidate screening applications; a preferred application involves its use in high-throughput screening, e.g. to identified "lead compounds."

10 A preferred embodiment of the invention relates to the use of the cDNA of SEQ ID NO:60 or part thereof, as a probe for examining individual aging at the gene expression level. Specifically, SEQ ID NO:60, or part thereof, can be used as a diagnostic reagent for various aging-associated diseases such as, but not limited to: arteriosclerosis, osteo-arthritis, dementia (including Alzheimer's disease) and Parkinson's disease.

15 In a related embodiment, the cDNA of SEQ ID NO:60 or part thereof, could be used to synthesize antisense oligonucleotides by methods well known to those skilled in the art. Antisense oligonucleotides can be used to inhibit the synthesis of the protein of SEQ ID NO:301, thereby preventing cell and tissue aging and/or promoting the rejuvenation of aged cells and tissues. These antisense oligonucleotides can also be used for in vivo or ex vivo treatment and prophylaxis of diseases caused by cellular senescence or aging-associated diseases such as, but not limited to: arteriosclerosis, osteo-arthritis, dementia (including Alzheimer's disease) and Parkinson's disease.

In a most preferred embodiment, SEQ ID NO:301, SEQ ID NO:60, or any part thereof, can be used as a pharmaceutical drug to treat pathologies such as, but not limited to cancers, inflammation, or infections. For example, when used as an antibacterial, antiviral and/or antifungal agent, inhibition of microbial proliferation could be achieved by either directly inhibiting microorganism growth (in the case of fungal and bacterial infections) or DNA synthesis of infected cells in the case of viral infections. The DNA of SEQ ID NO:60 or part thereof, can also be used to develop gene therapy products in the in vivo or ex vivo treatment of diseases and conditions such as cancers and inflammation. SEQ ID NO:301, SEQ ID NO:60, or any part thereof, may also be used: 1) as a probe for the diagnosis, 2) as a prophylaxis or 3) as a treatment of aging-associated diseases such as, but not limited to: arteriosclerosis, osteo-arthritis, dementia (including Alzheimer's disease) and Parkinson's disease. In a related embodiment, the protein of SEQ ID NO:301 or the cDNA of SEQ ID NO:60, or any part thereof, can be used in the development of drugs that will be used in the prophylaxis or treatment of the diseases stated above (e.g. diseases caused by cellular senescence, aging-associated diseases and diseases caused by cellular proliferation).

For the treatment or prevention of diseases and conditions associated with undesired proliferation, such as cancer, inflammation, or infection, the expression or activity of the present

protein can be increased using any of a number of methods. For example, polynucleotides encoding the protein can be introduced into the undesired cells, wherein the protein is then expressed and inhibits the further growth of the cells. In one such embodiment, the polynucleotides can be incorporated into liposomes comprising on their surface a specific molecule that directs the targeting of the liposome to a specific cell type (e.g. a tumor-specific antibody). Alternatively, the protein of SEQ ID NO:301 can itself be administered to the cells, e.g. as a fusion protein also comprising a specific targeting polypeptide moiety. Further, a compound that enhances the expression or activity of the protein can be administered to cells, preferably in a way that specifically targets the compound to undesired cells, e.g. chemically linked to a heterologous specific targeting molecule.

Protein of SEQ ID NO : 412 (internal designation 187-5-3-0-C7-CS)

The protein of SEQ ID NO : 412 encoded by the cDNA of SEQ ID NO : 171 is homologous to the human CDK4-binding protein p34^{SEI-1} (sptrembl accession number Q9UHV2). p34^{SEI-1} is a new CDK4 regulator that prevents p16INK4a from inhibiting the formation of cyclinD1-CDK4 complexes. p34^{SEI-1} seems to act as a growth factor sensor and may facilitate the formation and activation of cyclin D-CDK complexes in the face of inhibitory levels of INK4 proteins (Sugimoto et al., Genes Dev. 13:3027-3033 (1999)).

Progression through the cell cycle is a complex process that is regulated at many levels by several proteins. The activity of cyclin dependent kinases (CDK4 and CDK6) is regulated by the association of cyclin partner that acts as a positive effector and by two families of cdk inhibitors proteins (KIP) and the inhibitors of cdk4 (INK4) such as p16INK4a, which act as negative effectors (Sandhu et al., Cancer Detect. Prev.24:107-118 (2000)).

Cancer is a disease characterized by loss of cellular growth control, the molecular machinery of the cell cycle is involved in tumorigenesis. Many human tumors have been shown to have abnormality in this pathway resulting in either the functional inactivation of p16INK4a or the excessive activity of CDK4 (Palmero et al., Cancer Surv.27:351-357 (1996)).

It is believed that the protein SEQ ID No: 412 plays a role in the cell cycle regulation via the binding to a cyclin dependent kinase. Other preferred polypeptides of the invention are fragments of SEQ ID NO: 412 having any of the biological activity described herein. The binding activity of the protein of the invention or part thereof to a cyclin dependent kinase, as well as its role in cell cycle, may be assayed using any of the assays known to those skilled in the art including those described in Sugimoto et al., supra.

An embodiment of the present invention relates to methods of using the protein of the invention or part thereof to identify and/or quantify cyclin dependent kinases, preferably CDK4, in a biological sample, and thus used in assays and diagnostic kits for the quantification of such CDKs

in bodily fluids, in tissue samples, and in mammalian cell cultures. The binding activity of the protein of the invention or part thereof may be assessed using the assay described in Sugimoto et al., supra or any other method familiar to those skilled in the art. Preferably, a defined quantity of the protein of the invention or part thereof is added to the sample under conditions allowing the formation of a complex between the protein of the invention or part thereof and the cyclin dependent kinase to be identified and/or quantified. Then, the presence of the complex and/or the free protein of the invention or part thereof is assayed and eventually compared to a control using any of the techniques known by those skilled in the art.

In another embodiment, the invention relates to compositions and methods using the protein of the invention or part thereof to stimulate cell proliferation both in vitro and in vivo. For example, soluble forms of the protein of the invention or part thereof may be added to cell culture medium in an amount effective to stimulate cell proliferation.

The invention further relates to methods and compositions using the protein of the invention or part thereof to diagnose, prevent and/or treat several disorders associated with cell proliferation including but are not limited to, adenocarcinoma, sarcoma, lymphoma, leukemia, melanoma, myeloma, teratocarcinoma, cancers of the adrenal gland, bladder, bone, brain, breast, gastrointestinal tract, heart, kidney, liver, lung, ovary, pancreas, paraganglia, parathyroid, prostate, salivary gland, skin, spleen, testis, thyroid, uterus, and neurodegenerative disorders such as Alzheimer's disease (McShea et al., Am.J.Pathol.150(6):1933-1939 (1997)). For diagnostic purposes, quantification of the protein of the invention could be investigated, using Northern blotting, RT-PCR, immunoblotting and any of protocols known in the art, in biological samples and compared to the expression in control biological samples. Thus a diagnosis assay may be used, to determine altered expression of the protein of the invention, to correlate with diseases states and to evaluate the prognostic significance in diseases. For prevention and/or treatment purposes, inhibition of the endogenous expression of the protein of the invention using any of the antisense or triple helix methods described herein may be used. Alternatively, inhibitors for the protein's activity may be developed and use to inhibit and/or to reduce the protein's activity using any methods known to those skilled in the art. Antibodies which specifically bind to the protein of the invention may be generated using methods that are well known in the art and used as an antagonist.

30 Protein of SEQ ID NO : 299 (internal designation 184-1-4-0-C11-CS)

The protein of SEQ ID NO : 299 encoded by the cDNA of SEQ ID NO: 58 and found in fetal liver and liver, is orthologous to the Bola protein. The Bola family comprises the morpho-protein Bola from *E. coli* and its various homologs. The expression of Bola is growth rate regulated and is induced during the transition into the stationary phase. Bola is also induced by stress during early stages of growth and can have a general role in stress response. It has also been

suggested that BolA can induce the transcription of penicillin binding protein 6 and 5 (EMBO J. 1;8 (2) :3923-31 (1989)).

E. coli cells become thinner and shorter after a period of starvation or stationary-phase conditions; this altered morphology is an adaptative response of *E. coli* to general forms of stress.

- 5 The *bolA* gene seems to be involved in the switching between cell elongation and septation systems during the cell division cycle [J Bacteriol 170 :5169-5176 (1988)]. The regulation of *bolA* has been linked to the presence of gearbox promoter from which RNA is transcribed [Mol Microbiol 5 :2085-2091 (1991)].

- 10 Expression of *bolA* is governed by two promoters. P2 is located further upstream from the structural gene, is under the control of σ^d and transcribes *bolA* constitutively. The promoter P1, proximal to the structural gene, is a gearbox promoter under the control of σ^{54} from which *bolA* has been shown to be transcribed in an inverse growth rate-dependent fashion [J Bacteriol 173 :4474-4481 (1991)].

- 15 The alternate sigma factor σ^{54} is encoded by the gene *rpoS* and has been described as a central regulator for the induction of a set of specific genes involved in adaptation to stationary phase. It has, nevertheless, been shown that σ^{54} function is not confined to stationary phase. Significant increases in σ^{54} cellular levels were seen during exponential growth in response to forms of stress; genes under its control code for important adaptive regulators for general stress conditions [FEMS Microbiol Lett 30 :419-430 (1997)].

- 20 The smaller morphology caused by stress-induced overexpression of *bolA* reduces the surface area exposed to the environment and decreases the cell's surface-to-volume ratio.

- Identification of ortholog genes provides important information regarding functional and structural conservation within these orthologs throughout evolution. The concept of comparative gene identification has been previously used by many laboratories to search for orthologous genes
25 once a particular gene of interest has been identified in another species [Genome Res 10 (5) : 703-13 (2000)].

- The protein of invention contains a signal peptide corresponding to a short helix as predicted by software TopPred II [Clarosand and von Heijne, CABIOS applic. Notes, 10: 685-686 (1994)]. Thus, one aspect of this invention provides materials and methods for the delivery of
30 recombinant proteins to liver cells. The signal peptide, encoded by an appropriate polynucleotide, can be linked to another protein/polypeptide (also encoded by an appropriate polynucleotide). The recombinant gene, containing the signal peptide sequence, is expressed and the desired protein is delivered via the signal peptide. Methods of producing such gene fusions, the expression of such gene products, and their use are well known to the skilled artisan.

- 35 In another embodiment, BolA, or biologically active fragments thereof, can be used to modulate the stress response to environmental changes such as cytotoxic agents, heat shock, irradiation, genotoxic stress or growth factors.

In another aspect of the invention, SEQ ID NO: 299 is incorporated into a prokaryotic expression vector and transfected into prokaryotic cells unable to adapt to environmental stress or cells containing a BolA defect. The expression vector can, optionally, contain a promoter system such as that described supra (e.g., P1, P2, σ^d , σ^s , etc.) which typically controls bolA expression. The components necessary for transcription can be provided in one or more expression vectors. Prokaryotic cells thus transformed can be useful in bioremediation systems where environmental stress is commonly encountered. Thus, preferred prokaryotes for the practice of this aspect of the invention lack bolA, or contain a bolA defect, and are known to be useful for bioremediation.

In another embodiment, the subject invention provides methods and compositions to selectively identify liver tissues. The protein encoded by SEQ ID NO: 299 can be used to synthesize specific polyclonal or monoclonal antibodies using any techniques known to those skilled in the art. These antibodies can be used to selectively identify liver tissue according to well-known histological immunoassays. The ability to immunologically identify tissue samples is industrially important for analysis of mismarked biopsy samples (e.g., laboratory errors) where the origin of the tissue sample is in question, or simply to verify that a tissue sample originated from liver. The antibodies can also be used to identify cancer metastases originating from the liver.

Further, antibodies provided by the subject invention can also be used to assay animal feeds for the presence of liver or liver by-products. As is known, many animal feeds contain animal protein. The use of animal feeds containing animal protein has been associated with disorders in both animals and humans (the most notorious of which is bovine spongiform encephalitis). This has resulted in the banning of animal protein in feeds provided to animals. However, to ensure compliance with such bans, animal feeds must be tested. Thus, the antibodies of the invention can be used to test animal feeds for the presence of liver according to methods known to those skilled in the art.

25 Proteins of SEQ ID NOs: 249 and 288 (internal designation 105-037-2-0-H11-CS and 174-7-4-0-H1-CS respectively).

The 403-amino-acid-long protein of SEQ ID NO: 249 encoded by the cDNA of SEQ ID NO: 8 is extensively homologous to the protein of SEQ ID NO: 288 encoded by the cDNA of SEQ ID NO: 47 with the exception of five amino acids in positions 192-194, and 298-299 which are not present in protein of SEQ ID NO: 288. It is likely that the two proteins are the result of an alternative splicing and display similar functions and utilities.

The 403-amino-acid-long protein of SEQ ID NO: 249, overexpressed in salivary glands, exhibits extensive homology to the mus musculus hypothetical protein (Genbank accession number AB030196). The amino acid residues of protein of SEQ ID: 249 show a high degree of identity to the Genbank sequence. However, the protein of Genbank sequence does not have the twenty amino acids (192 to 194, and 298-303, 353-354, 380-381, and 387-393) and also displays 35 different

amino acids from the SEQ ID NO: 249. In addition, four transmembrane domains are predicted for the protein of SEQ ID NO: 249 from positions 31 to 51, 75 to 95, 154 to 174, and from 236 to 256 as predicted by the software TopPred II (Claros and von Heijne, *CABIOS applic. Notes*, 10 : 685-686 (1994)).

5 When expressed in *E. Coli*, the matched sequence suppresses bacterial growth (Inoue et al, *Biochem Biophys Res Commun* 268:553-61 (2000)). It is therefore believed that the proteins of SEQ ID NO: 249 and 288 or a bacterial growth suppressing fragment thereof can be used to suppress bacterial growth by contacting bacteria (gram negative or gram positive) with the polypeptides of the invention. The growth inhibiting activity of the protein of the invention or part
10 thereof may be assayed using any of the assays known to those skilled in the art including those described in Inoue et al, supra.

 In accordance with one aspect of the invention, methods and compositions using the protein of the invention or a fragment thereof to suppress bacterial growth are provided. In a preferred embodiment, the protein of the invention is expressed in a bacteria, preferably *E. coli*, using
15 recombinant DNA technology methods known to those skilled in the art. The expressed protein can then be used to inhibit bacterial growth. The effects of the expressed protein and analogs or antagonists thereof can be assessed using any methods or techniques known to those skilled in the art.

 Further included in the invention are the polypeptides encoded by the human cDNA of
20 clone 105-037-2-0-H11-CS-SD. The polypeptides of SEQ ID NO: 249 may be interchanged with the corresponding polypeptides encoded by the human cDNA of clone 105-037-2-0-H11-CS-SD. Further included in the invention are polynucleotides encoding said polypeptides. Preferred polynucleotides are those of SEQ ID NO: 8 and of the human cDNA of clone 105-037-2-0-H11-CS-SD.

25 Nucleotide sequences encoding the polypeptides of SEQ. ID. NOs: 249 and 288 can be used to generate probes for the detection of related genes. Vectors expressing the nucleotide sequence can be used can be used to express the polypeptide in target cells. Antisense nucleotides can be used to inhibit the expression of the polypeptide.

 Thus, in another embodiment of the invention the protein or a fragment thereof can be used
30 as a marker protein to selectively identify tissues, preferably salivary glands. For example, the protein of the invention or a fragment may be used to synthesize specific antibodies using any techniques known to those skilled in the art. Such tissue-specific antibodies may then be used to identify tissues of unknown origin, for example, forensic samples, differentiated tumor tissue that has metastasized to foreign bodily sites, or to differentiate different tissue types in a tissue cross-
35 section using immunochemistry. In another embodiment, polynucleotides encoding the protein of SEQ. ID. NO. 249 can be used for *in situ* hybridization.

The transcript coding for the protein gng3lg (Genebank accession number AF069954) is transcribed from a bidirectional promoter divergently with the transcript coding for the gamma 3 subunit protein called gng3, a novel human G binding protein gamma-3 (HGPG) (Genbank accession number AF069953) and this organization is conserved across species within the human
5 genome (*Downes et al, Genomics, 53:220-230 (1998)*).

Several genes which are linked in common physiological functions share a common divergently bidirectional promoter like α B crystallin and α crystallin, collagen type IV A1 and A2, surf1-3, and surf 5 genes, dihydrofolate reductase and 2 mismatch repair1 (*Iwaki et al., Genomics, 45: 386-394 (1997)*, *Burbelo et al., Proc. Natl. Acad. Sci. USA 85: 9679-9682 (1988)*, *Kaytes et al., J. Biol. Chem, 263: 19274-19277 (1988)*, *Poschl et al., EMBO J., 7:2687-2695 (1988)*, *Soininen et al., J. Biol. Chem, 263: 17217-17220 (1988)*, *Garson et al., Genomics, 30: 163-170 (1995)*, *Williams et al., Mol. Cell. Biol., 6: 4558-4569 (1986)*, *Fujii et al., J. Biol. Chem., 264: 10057-10064 (1989)*). The heterodimeric G proteins, a family of GTPases are present in all cells and control a variety of functions (metabolic, humoral, neural and developmental) by transducing hormonal,
15 neurotransmitter and sensory signals into an array of cellular responses. Triggered by cell surface receptors, each G protein regulate the activity of a specific effector including adenylate cyclase, phospholipase C, and ion channels protein which initiate appropriate biochemical responses. In view of this, it is believed that the transcript coding for the proteins of SEQ ID NO: 249 shares common regulatory elements with gng3 gene and that the products of such genes which are protein
20 of SEQ ID NO: 249 and gng3 are physiologically coupled in unknown ways. Thus, in an embodiment of the invention, the protein of SEQ ID NO: 249 or part thereof may be used to regulate signal transduction of hormonal, neurotransmitter, and sensory signals to provide an array of cellular responses.

In yet another embodiment of the invention, the polypeptides of the present invention and
25 the related polynucleotides may be used to treat several types of disorders including, but not limited to, cancer, neurodegenerative diseases, cardiovascular disorders, hypertension, renal injury and repair, septic shock.

In one embodiment, the protein of SEQ. ID. NO: 249 or a fragment, derivative or analog thereof may be administered to a subject to treat or prevent a disorder associated with decreased
30 expression or activity of the protein.

In a further embodiment, a vector capable of expressing the protein of SEQ. ID. NO: 249 may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of the protein. Naked nucleotides encoding the protein of SEQ. ID. NO 249 may also be used.

35 In yet another embodiment, a pharmaceutical composition comprising a substantially purified the protein of SEQ ID NO: 249 may be administered to a subject to treat or prevent a disorder associated with decreased expression of the protein.

In still another embodiment, the polypeptide of SEQ ID NO: 249 can be used to develop and screen antagonists. For example, purified polypeptide can be used to develop antibodies or to screen libraries of pharmaceutical agents to identify those that inhibit the physiological functions of the protein.

5 Thus, in a further embodiment, an antagonist of the protein of SEQ ID NO: 249 can be administered to a subject to prevent or treat a disease associated with increased expression or activity of the protein. Similarly, in another embodiment a vector expressing the complement of a polynucleotide sequence encoding the protein of SEQ ID NO: 249 may be administered to decrease the expression of the protein.

10 The protein of SEQ ID NO: 249 displays a leucine zipper pattern situated near its NH₂ terminal part (position 20 to 41). Thus, it is believed that the protein of SEQ ID NO: 249 is able to dimerize either with itself (homo-dimerisation) or with an heterologous protein (hetero-dimerisation) of interest, through the mediation of its leucine zipper domain. Preferred polypeptides of the invention are polypeptides comprising leucine zipper domains fragments and fragments
15 having any of the biological activities described herein. The multimerization activity of the protein of the invention or part thereof may be assayed using any of the assays known to those skilled in the art including circular dichroism spectrum and thermal melting analyses as described in US patent 5,942,433. The utilities of proteins containing leucine zipper domains, such as the protein of SEQ ID No: 249, are described elsewhere in the application.

20 Protein of SEQ ID NO: 259 (internal designation 114-016-1-0-H8-CS)

The protein of SEQ ID NO: 259, herein referred to as HOPP, is encoded by clone 114-016-1-0-H8-CS (SEQ ID NO: 18). This protein is homologous to a protein of *Arabidopsis thaliana* (ASY1) and *Saccharomyces cerevisiae*, (HOP1) (Caryl A.P. et al. Chromosoma, 109, 62-71; Hollingsworth N.M. et al. Cell, 61, 73-84).

25 In addition, the 394-amino-acid protein of SEQ ID NO: 259 displays a pfam HORMA domain from position 22 to 230. The HORMA domain is a common structural element in mitotic checkpoints, chromosome synapsis and DNA repair. For example, the HORMA domain was found in: (1) MAD2, a key component of the mitotic-spindle-assembly checkpoint (reviewed in Straiht AF. Current Biology 1997, 7:613-616); (2) HOP1, a conserved protein that is involved in meiotic-
30 synaptonemal-complex assembly (Hollingsworth N.M. et al. Cell, 61, 73-84); and (3) in Rev7p, a subunit of the yeast DNA polymerase "epsilon" that is involved in translation, template independent DNA synthesis (Aravind L. and Koonin E.V., Trends Biochem Sci. 1998 Aug;23(8):284-6).

The pairing of homologous chromosomes during meiotic prophase culminates in the formation of the synaptonemal complex (SC), which is a ribbon-like, proteinaceous structure that
35 holds homologous chromosomes in close apposition along their entire length. The synaptonemal complex (SC) is a prominent and evolutionally well conserved structure which is strictly meiotic.

Evidence from mutant phenotypes supports the hypothesis that recombination and SC formation are mutually interdependent processes. First, although not required for homology recognition, the SC could promote interhomolog interactions in situations where the normal processes have failed (*e.g.*, interlocking, heterologous pairing, etc.). Second, polymerization of the SC components might
5 permit the recombination process to progress by modulating the number and localization of reciprocal versus exchanges (*i.e.* interference). Third, the SC may play an important role in meiotic chromosome structure and especially inter-sister interactions.

Synapsis of homologous chromosomes is a key event in meiosis as it is essential for normal chromosome segregation and is implicated in the regulation of crossover frequency (for review see
10 Zickler D., J Soc Biol 1999;193(1):17-22). Mutants in HOP1 and ASY1, both proteins having significant homology to the protein of SEQ ID NO: 259, display decreased levels of meiotic crossover and intragenic recombination between markers on homologous chromosomes (Hollingsworth N.M., Byers B., Genetics 1989 Mar;121(3):445-62 ; Caryl AP et al. Chromosoma 2000;109(1-2):62-71).

15 Thus, the invention relates to methods and compositions using the protein encoded by clone 114-016-1-0-H8-CS or polynucleotide of SEQ ID NO: 18, or biologically active fragments thereof, to restore normal chromosome segregation in cells by administration of compositions comprising HOPP polypeptide, or polynucleotide encoding a HOPP polypeptide, encoded by clone 114-016-1-0-H8-CS, or polynucleotide in therapeutically effective amounts. The loss of normal chromosome
20 segregation in normal cells leads to aberrant chromosome segregation events, a hallmark of tumor progression. HOPP proteins, encoded by clone 114-016-1-0-H8-CS, can be targeted to the nucleus by nuclear targeting sequences according to well-known methods. Nuclear targeting sequences (or NLS) can be chemically or recombinantly attached to HOPP. Alternatively, the HOPP gene can be used in known gene therapy protocols to restore normal chromosome function.

25 Infertility due to gametogenic failure is frequently associated with structural autosomal abnormalities. Recent meiotic studies, at pachytene stage, have shown a failure around the breakpoints, an association of the translocation figure with the sex chromosomes, and the frequent involvement of the acrocentric chromosomes. Two main models are proposed to explain the male sterilizing effect of rearrangements. The impairment of spermatogenesis could be the result of: 1)
30 the XY-autosome interaction; or 2) the disruption around the breakpoints at the pachytene stage. These defects may contribute significantly to germ-cell atresia (for review see Luciani JM, Guichaoua MR Reprod Nutr Dev 1990; Suppl 1:95s-103s and Miklos GL. Cytogenet Cell Genet. 1974;13(6):558-77). Thus the subject invention also relates to methods and compositions of using the protein of SEQ ID NO: HOPP or clone 114-016-1-0-H8-CS, or biologically active fragments
35 thereof, to reduce the incidence of infertility due to gametogenic failure. HOPP can be introduced into sperm as described in the preceding paragraphs. HOPP, optionally joined to a NLS sequence;

can also be introduced into sperm or eggs by other methods well known in the art (such as electroporation or microinjection).

The protein of SEQ ID NO: 259, encoded by clone 114-016-1-0-H8-CS, can also arrest cell division in human cells if the mitotic spindle apparatus is improperly attached to the chromosomes (Allshire R.C. Current Opinion in Genetics and Development 1997, 7:264-273). In the absence of functional protein of SEQ ID NO: 259, cells exposed to drugs which inhibit the formation of a mitotic spindle, such as benomyl, vinblastine, nocodazole, etc. would be expected to undergo rapid cell death due to massive chromosome loss. Human cells containing HOPP would be expected to survive such drug treatment because they are able to stop dividing prior to the chromosome loss event. Tumor cells that are hypersensitive to chemotherapeutic agents, which inhibit the formation of the mitotic spindle, may be sensitive to these drugs because they are defective in the checkpoint protein. Thus, screening assays for the presence or absence of the protein in a given tumor would provide an indication of the chemosensitivity of a particular tumor. The present invention therefore includes methods of determining prognostic benefit of treating a patient with a chemotherapeutic agent or determining which chemotherapeutic agent from a group of at least two would a patient more likely benefit from. Furthermore, the loss of checkpoint function in a normal cell may predispose that cell to aberrant chromosome segregation events, a hallmark of tumor progression. Thus the antibodies, polypeptides and polynucleotides of the present invention would be useful in diagnosing particular cancers.

Polyclonal antibodies can be produced by injecting a host animal such as rabbit, rat, goat, mouse or other animal with an immunogen of this invention. The sera is extracted from the host animal and is screened to obtain polyclonal antibodies which are specific to the immunogen. Methods of screening for polyclonal antibodies are well known to those of ordinary skill in the art such as those disclosed in Harlow & Lane, Antibodies: A Laboratory Manual, (Cold Spring Harbor Laboratories, Cold Spring Harbor, N.Y.: 1988) the contents of which are hereby incorporated by reference.

The monoclonal antibodies can be produced by immunizing, for example, mice with an immunogen according to the invention. Methods of producing monoclonal antibodies are well-known in the art and include those methods Kohler, B. and Milstein, C., Nature (1975) 256: 495-497. Hybridomas can be expanded, if desired, and supernatants can be assayed by conventional immunoassay procedures, for example radioimmunoassay. Positive clones can be further characterized. Hybridomas that produce the desired antibodies can be grown in vitro or in vivo using known procedures. The monoclonal antibodies can be isolated by conventional immunoglobulin purification procedures such as ammonium sulfate precipitation, gel electrophoresis, dialysis, affinity chromatography, and ultrafiltration.

Antibodies of the invention can be labeled with a detectable moiety. As noted above, a "detectable moiety" is well known to those of ordinary skill in the art and include, but are not

limited to, a fluorescent label, a radioactive atom, a paramagnetic ion, biotin, a chemiluminescent label or a label which can be detected through a secondary enzymatic or binding step.

The invention further provides a method of determining the susceptibility of a tumor sample to treatment with a mitotic spindle inhibitor which comprises steps of: a) contacting the tumor sample with an antibody, wherein the antibody is labeled with a detectable moiety and is capable of specifically binding to the protein of invention, or a fragment thereof, and b) assaying for the presence of an immunocomplex formed in step (a). The absence of the immunocomplex indicates that the tumor would be susceptible to treatment with a mitotic spindle inhibitor such as benomyl, vinblastine, and nocodazole.

10 The subject invention also provides a pharmaceutical composition comprising nucleic acid encoding the protein of SEQ ID NO: 259 and a carrier. In one aspect of this invention, the compositions are capable of passing through a cell membrane and provide for the expression of the protein of invention. As used herein, the term "carrier" includes pharmaceutically acceptable carriers and encompasses any of the standard pharmaceutically accepted carriers, such as phosphate buffered saline solution, water, emulsions such as an oil/water emulsion or a triglyceride emulsion, various types of wetting agents, tablets, coated tablets and capsules. carriers contain excipients such as starch, milk, sugar, certain types of clay, gelatin, stensic acid, talc, vegetable fats or oils, gums, glycols, or other known excipients. Flavor and color additives or other ingredients can also be included. In addition to the standard characteristics of the pharmaceutically acceptable carriers, the

15 "suitable" carriers of the subject can also include those carriers which are able to penetrate the cell membrane. Therefore in one embodiment of the pharmaceutical composition the pharmaceutically acceptable carrier binds to a receptor on a cell capable of being taken up by the cell after binding to the structure.

20 This invention further provides a method of suppressing tumor formation in a subject which comprises administering a nucleic acid encoding the protein of invention in an amount effective to enhance expression of this protein.

Proteins of SEQ ID NOs: 311 and 312 (internal designation 188-28-4-0-B12-CS.corr and 188-28-4-0-B12-CS.fr respectively)

The 466-amino-acid-long protein of SEQ ID NO: 311, encoded by the human cDNA of clone 188-28-4-0-B12-CS or the cDNA nucleotide sequence of SEQ ID NO: 70, is related to proliferating-cell nucleolar antigen p120 (Genbank accession number M32110) encoded by noll; and the yeast nucleolar protein Nop2p coded by nop2 (Genbank accession number U12141). SEQ ID NO: 311 (encoded by clone 188-28-4-0-B12-CS.corr) shows strong homology with three proteins described as homologs of p120 (Genbank accession number AK002229 and Geneseqp

30 accession numbers: Y86441, Y86442). In addition, the protein of SEQ ID NO: 311 is a polymorphic variant of SEQ ID No: 312 encoded by the cDNA of SEQ ID No: 71.

In addition, the protein of the invention exhibits the pfam NOL1/NOP2/sun family signature from positions 201 to 276. This motif is also found by motif from positions 230 to 245. The NOL1/NOP2/sun family include p120 and Nop2p. These proteins are involved in nucleolar structure and activity as well as the regulation of cell cycle.

- 5 Freeman J.W. et al. (Cancer Res. 48: 1244-51, 1988) identified p120, a 120-kD nucleolar antigen associated with proliferating cells. This protein is a proliferation-associated antigen that is temporally regulated during the cell cycle and demonstrates a dramatic increase in expression at the G1-S boundary. This suggests that p120 can play a role in the regulation of the cell cycle and the increased nucleolar activity that is associated with cell proliferation (Fonagy A. et al. (1993) J. Cell. 10 Physiol. 154:16-27).

The human p120 protein is also the most cancer specific of the identified proliferation-associated nucleolar proteins. Antigen p120 was detectable in a broad range of human malignant tumors but not in benign tumors or corresponding normal tissues. The antigen was not detectable in growth-arrested cells but was expressed early in G1 of the cell cycle.

- 15 Overexpression of human p120 leads to the transformation of NIH 3T3 cells. Expression of antisense p120 constructs causes the p120-transformed cells to revert to their original phenotype. Perlaky L. et al. [Cancer Res. 52: 428-36 (1992)] and Valdez et al. [Cancer Res. 52: 5681-87 (1992)] reported that the middle region of antisense p120 RNA inhibited proliferation of NIH 3T3 cells to approximately the same extent as the full-length antisense construct. The predicted mouse 20 and human P120 proteins are 63% identical.

- Another protein of the Nol1/Nop2/Sun family, Nop2p, coded by the gene NOP2, has a role in nucleolar function during the onset of growth and in the maintenance of nucleolar structure (de Beus et al. (1994) J. Cell Biol. 127:1799-813). The two proteins, p120 and Nop2p, are associated to ribosomal RNA in pre-ribosomal particles and can mediate the maturation process of the ribosome 25 (Hong B. et al. (1997) Mol. Cell Biol. 17:378-88; Gustafson W.C. et al. (1998) Biochem. J. 331:387-93).

- The subject invention provides the polypeptides encoded by the human cDNA of clone 188-28-4-0-B12-CS and polynucleotide sequences encoding the same amino acid sequences. Also included in the invention are biologically active fragments of the protein encoded by the human 30 cDNA of clone 188-28-4-0-B12-CS and polynucleotide sequences encoding these biologically active fragments. "Biologically active fragments" are defined as those peptide or polypeptide fragments having at least one of the biological functions of the full length protein (e.g., the ability to transform cell lines in vitro.).

- The invention also provides variants of the protein of SEQ ID NO: 311, encoded by clone 35 188-28-4-0-B12-CS. These variants have at least about 80%, more preferably at least about 90%, and most preferably at least about 95% amino acid sequence identity to the amino acid sequence encoded by clone 188-28-4-0-B12-CS. Variants according to the subject invention also have at

least one functional or structural characteristic of the protein encoded by clone 188-28-4-0-B12-CS. The invention also provides biologically active fragments of the variant proteins. Unless otherwise indicated, the methods disclosed herein can be practiced utilizing the protein encoded by clone 188-28-4-0-B12-CS, or clone 188-28-4-0-B12-CS, or variants thereof. Likewise, the methods of the
5 subject invention can be practiced using biologically active fragments of the protein encoded by clone 188-28-4-0-B12-CS, clone 188-28-4-0-B12-CS, or variants of said biologically active fragments.

Because of the redundancy of the genetic code, a variety of different DNA sequences can encode the amino acid sequence provided by clone 188-28-4-0-B12-CS. It is well within the skill
10 of a person trained in the art to create these alternative DNA sequences encoding proteins having the same, or essentially the same, amino acid sequence. These variant DNA sequences are, thus, within the scope of the subject invention. As used herein, reference to "essentially the same" sequence refers to sequences that have amino acid substitutions, deletions, additions, or insertions that do not materially affect biological activity. Fragments retaining one or more characteristic
15 biological activity of the protein encoded by clone 188-28-4-0-B12-CS are also included in this definition.

"Recombinant nucleotide variants" are alternate polynucleotides which encode a particular protein. They can be synthesized, for example, by making use of the "redundancy" in the genetic code. Various codon substitutions, such as the silent changes which produce specific restriction
20 sites or codon usage-specific mutations, can be introduced to optimize cloning into a plasmid or viral vector or expression in a particular prokaryotic or eukaryotic host system, respectively.

In one aspect of the subject invention, SEQ ID NO: 311, encoded by clone 188-28-4-0-B12-CS, and variants thereof, can be used to generate polyclonal or monoclonal antibodies. Both biologically active and immunogenic fragments of the amino acid sequence or variant proteins can
25 be used to produce antibodies. Polyclonal and/or monoclonal antibodies can be made according to methods well known to the skilled artisan. Antibodies produced in accordance with subject invention can be used in a variety of detection assays known to those skilled in the art. Another aspect of this invention provides monoclonal and polyclonal antibodies which do not cross-react with known p120 proteins.

30 In one embodiment, the protein encoded by clone 188-28-4-0-B12-CS, variants of said protein, and biologically active fragments of the protein or said variants can be used as a nucleolar-fraction marker in nuclear fractionation studies or as a marker of pre-ribosomal particles, in methods well known to the skilled artisan.

In another embodiment, the protein encoded by clone 188-28-4-0-B12-CS, variants of said
35 proteins, and biologically active fragments thereof, can be used as a proliferation marker in neoplastic cells. Alternatively, quantitative immunoassays can be used to assess the levels of the protein in resected cancerous and normal tissues. Alternatively, levels of the protein encoded by

clone 188-28-4-0-B12-CS can be compared between an individual and a "normal" control group as a prognostic indicator of malignancy. Further, the relationship between protein expression and cell proliferation can be assayed using cancer cell lines. Thus, the protein of the invention or part thereof can be used as a marker for proliferation in human cancer cells in vivo and in vitro. If the absence of expression of the protein of the invention on normal and benign tumors is confirmed, it could serve as a marker of malignant cancer cells. The proliferation rate of cancer cells can be also determined by quantitative analysis of the expression of the protein encoded by 188-28-4-0-B12-CS, or biologically active fragments thereof, according to methods described in Trere D. et al. (J. Pathol. 192:216-20, 2000).

10 The transforming activity of the protein of the invention can be assayed as described in Perlaky et al. (Anticancer Drug Des. 8:3-14, 1993). Thus, polynucleotides encoding the (188-28-4-0-B12-CS) protein can be used to induce transformation on NIH/3T3 cells in vitro. Alternatively, the polynucleotide encoding (188-28-4-0-B12-CS) can be used to provide antisense oligonucleotides useful in antisense therapeutic protocols according to methods known in the art.

15 Protein of SEQ ID NO: 406 (internal designation 174-32-4-0-F8-CS)

The 378-amino-acid-long protein of SEQ ID NO: 406 encoded by the cDNA of SEQ ID NO: 165 is expressed in tissues such as colon, prostate and salivary glands and overexpressed in colon and prostate. The C-terminus of the protein of the invention is homologous to the human retinoblastoma-binding protein, RbAp48 (Qian YW et al. (1993) Nature 364:648-652, GenBank accession number: X74262) and to its homologues conserved in other organisms including mouse (GenBank accession number: Q60972) and C. elegans (GenBank accession number: AF116530). The protein of the invention contains also two internal WD-repeat clusters (Prosite PS00678, amino acid positions 267 to 304 and positions 333 to 370, respectively), a structural motif involved in proteins interaction in signal transduction pathway and transcription regulation (Neer EJ et al. (1994) Nature 371:297-300; Neer EJ et al (1996) Cell 84:175-178).

The retinoblastoma protein (Rb) is the product of the retinoblastoma gene. Deletion or inactivation of both Rb alleles is essential in the formation of human retinoblastoma in both hereditary and sporadic forms (Benedict WF et al. (1983) Science 219: 973-975).

Loss-of-function mutations in the Rb gene is also found in many other tumor types, including osteosarcoma, breast carcinoma, small cell lung carcinoma, bladder carcinoma, prostate carcinoma and soft tissue sarcoma (Bookstein R et al. (1991) Crit. Rev. Oncog. 2:211-227). Introduction of the wild-type Rb gene into cultured retinoblastoma cells suppresses cells growth and their tumorigenicity in nude mice (Huang HJ et al (1988) 242:1536-1566). Expression of normal Rb protein in prostate carcinoma, osteosarcoma, breast carcinoma and bladder carcinoma cells also suppresses their neoplastic phenotype, thus establishing the Rb gene as a tumor suppressor (Reviewed by Weinberg RA (1991) Science 254:1138-1146).

It has been shown that the Rb gene product is a nuclear phosphoprotein that undergoes cyclic phosphorylation and dephosphorylation during cell cycling. Rb is unphosphorylated or "underphosphorylated" during early G1 phase, and become phosphorylated just before S phase, and remains phosphorylated until late mitosis. Injection of unphosphorylated Rb protein into cell during early G1 phase inhibits the entry into S phase, suggesting that some of the growth suppressor functions of Rb may be carried out by the underphosphorylated form of Rb (Goodrich et al. (1991) Cell 67:293-302; Hinds PW et al. (1992) Cell 70:993-1006). Rb protein not only regulates cell cycle, but is also involved in cell differentiation. For example, lens epithelial cells in Rb-deficient mouse fail to terminally differentiate and undergo apoptosis (Morgenbesser et al.(1994) Nature 371:72-74).

It has been demonstrated that Rb protein inhibits cellular growth and proliferation through interactions with multiple cellular proteins that interfere with these cellular protein's downstream actions. For example, Rb protein is able to form specific complexes with transcriptional factor E2F, which regulates the expression of a set of genes essential for the G1 to S phase transition (Nevin JR et al (1992) Science 258:424-429). The Rb protein restrains cell cycle progression by masking the E2F transactivation domain and by blocking the interaction of surrounding enhancer elements and basal transcription complex (Weintraub SJ et al (1995) Nature 375:812-815). Association of Rb and UBF, a ribosomal transcription factor, results in suppression of the synthesis of ribosomal RNA by RNA polymerase I (Cavanaugh LI et al (1995) Nature 374:177-180; Mancini M et al (1994) Proc.Natl.Acad.Sci.USA 91:418-422).

RbAp48 was first identified as a major protein from Hela cell that binds to a putative functional domain at the C-terminus of the Rb protein. Only unphosphorylated and hypophosphorylated forms of the Rb protein were coprecipitated with RbAp48. Like Rb protein, RbAp48 is a ubiquitously expressed nuclear protein that shares sequence homology with MS11, a negative regulator of the Ras-cAMP pathway in the yeast *Saccharomyces cerevisiae*. Overexpression of RbAp48 can convert the yeast mutant strains from heat-shock sensitivity to heat-shock resistant, similar to the result obtained from MS11 overexpression. Thus, the human RbAp48 is a functional homologue of MS11 (Qian YW et al. (1993) Nature 364:648-652).

Rbap48 protein was later found to be the p48 subunit of mammalian chromatin assembly factor 1 (CAF-1) and to be present in histone deacetylase complex (Parthum MR et al (1996) Cell 87:85-94). CAF-1 from human cell nuclei consists of three subunits of p150, p60 and p48 and is involved in assembling of histone3 and histone4 onto nascent nucleosome structure during DNA replication in S phase (Kaufman FD et al (1995) Cell 81:1105-1114). Indeed, some transcriptional repressors function through the recruitment of the histone deacetylase complex (HDAC), the latter acts by acetylating or deacetylating the tail protruding from the core histones, thereby modulating the local structure of chromatin (Reviewed by Pazin MJ et al (1997) Cell 89:325-328). Rb protein recruits HDAC for binding to E2F to repress transcription (Brehm A et al (1998) Nature 391:597-601; Magnaghi-Jaulin L

et al (1998) Nature 391:601-604). It was also reported that the p48 subunit of chicken CAF-1 can bind to chicken HDAC in vitro through interaction of WD-40 repeats presented in both protein sequences (Ahmad A et al (1999) J.Biol.Chem. 274:16646-16653).

The WD-40 protein family is characterized by the repetition of a loosely conserved repeat of approximately 40 amino acids, each repeat being separated from each other by a Trp-Asp dipeptide sequence. The conserved core of this repeat, which usually ends with the amino acids Trp-Asp (WD), was first identified in the beta-subunit of the heterotrimeric GTP-binding protein, G-protein (Fong H et al. (1986) Proc.Natl.Acad.Sci.USA 83:2162-2166). Among the WD-40 proteins identified to date, none are enzymes, and all seem to have regulatory functions (Neer, E. J. et al. (1994) Nature 371:297-300).

10 A number of WD repeat proteins have been localized to the nucleus and function in the repression of transcription. These include Tup1, Hir1, and Met30 in *S. cerevisiae*; SCON2 in *Neurospora crassa*; extra sex combs and Groucho in *Drosophila*; COP1 in *Arabidopsis thaliana*; and HIRA and the family of TLE proteins in humans. These WD-40 proteins turn off a wide variety of genes, including those involved in segmentation, sex determination, and neurogenesis (controlled by Groucho) and those

15 involved in photomorphogenesis (controlled by COP1). All of these WD-40 containing proteins have been proposed to fold into propellers in which the internal beta-strands form a rigid skeleton that is fleshed out on the surface by specialized loops to which other proteins bind (Lambright DG et al (1996) Nature 379:311-319; Sondex J et al (1996) Nature 379:369-374).

Thus, discovery of new Rb-binding proteins is necessary to design methods of regulating cell growth and block tumorigenesis through the control of tumor suppressor proteins in their interaction with oncogene products and may provide new compositions which are useful in the diagnosis, prevention and treatment of cancer and developmental disorders.

It is believed that the protein of SEQ ID NO: 406 or part thereof plays a role in the control of gene expression, probably as a transcription repressor. The protein of the invention is thought to be able to bind to other proteins, preferably to nuclear proteins, more preferably to Rb. Preferred polypeptides of the invention are polypeptides comprising fragments of SEQ ID NO: 406 from position 159-373, 267-304 and 333-370. Other preferred polypeptides of the invention are polypeptides comprising fragments of SEQ ID NO: 406 having any of the biological activity described herein. The ability of the protein of the invention or part thereof to function as a transcription repressor may be assessed using techniques well known to those skilled in the art including those described previously (Weintraub SJ (1995) Nature 375:812-815; Qian YW (1995) J.Biol.Chem. 270:25507-25513). The ability of the protein of the invention or part thereof, especially fragments containing WD-repeats, to bind to other proteins may be assessed using techniques well known to those skilled in the art including those described herein. For example, the protein of the invention could be used as a "bait" protein in a yeast double hybridization system (e.g. Gal-4-based system from Clontech) to isolate and eventually to identify its interacting protein partner in vivo from a cDNA library. Alternatively, the protein of the invention or part thereof can be used either in a pure form or in a fusion form (linked to a reporter gene

30

35

product, such as alkaline phosphatase) to screen a phage cDNA expression library derived from selected tissues or cell types of a given organism (Scott et al (1990) Science 249:386-390; Lam et al (1992) Nature 354:82-84). Preferably, the binding ability of protein of the invention is tested in mammalian cell transfection experiments. When fused in-frame to a suitable peptide tag in expression
5 vector, such as [His]₆ in the pRset expression plasmid vector (Invitrogen) and introduced into culture cells, the proteins that bind to the expressed fusion protein can be immunoprecipitated using anti-[His]₆ antibody. This approach can also be employed to confirm the findings obtained from either yeast double hybridization system or in vitro phage peptide library screening. In this case, the putative interacting partner protein will be fused to a distinct tag in a second expression vector and co-
10 transfected into culture cells. True binding complex will be co-immunoprecipitated with the two different anti-tag antibodies. In a particular embodiment, an affinity chromatography method is carried out to identify the interacting protein partners in vitro from cell lysates as performed for the identification of the RbAp48 protein (Qian YW et al. (1993) Nature 364:648-652).

An embodiment of the present invention relates to methods of using the protein of the
15 invention or part thereof, particularly polypeptides containing WD-motifs, or derivative thereof to identify and/or quantify binding proteins, preferably nuclear proteins, more preferably Rb, in a biological sample, and thus used in assays and diagnostic kits for the quantification of such binding proteins in bodily fluids, in tissue samples, and in mammalian cell cultures. Such assays may be particularly useful as diagnostic or prognostic tools in the detection and monitoring of a disorder
20 linked to dysregulation of expression of a transcription regulator. Such assays may thus be very useful to assess the level of the tumor suppressor Rb in disorders including but not limited to developmental disorders, cancers such as retinoblastoma, prostate carcinoma, osteosarcoma, breast carcinoma and bladder carcinoma. The binding activity of the protein of the invention or part thereof may be assessed using any method familiar to those skilled in the art. Preferably, a defined
25 quantity of the protein of the invention or part thereof is added to the sample under conditions allowing the formation of a complex between the protein of the invention or part thereof and the binding protein to be identified and/or quantified. Then, the presence of the complex and/or the free protein of the invention or part thereof is assayed and eventually compared to a control using any of the techniques known by those skilled in the art.

30 Another embodiment of the present invention relates to compositions and methods using the protein of the invention or part thereof or derivative thereof to block gene transcription either in vitro or in vivo. In a preferred embodiment, the protein of the invention or part thereof or derivative thereof is added in an effective amount to an in vitro culture to inhibit gene expression and thus cell proliferation using molecular biology techniques known to those skilled in the art allowing the import of the protein
35 from the extracellular medium to the cell's nucleus. In another embodiment, eukaryotic cells are genetically engineered in order to express the protein of the invention or part thereof under specific

conditions in order to prevent further proliferation of such cells upon demand such as infection, transformation, activation, differentiation, end of a production process.

A preferred embodiment of the invention relates to compositions or methods using SEQ ID NO: 406, SEQ ID NO: 165 or part thereof to diagnose, treat and/or prevent disorders including but not limited to disorders linked to dysregulation of gene transcription such as cancers and other disorders relating to abnormal cellular differentiation, proliferation, or degeneration, including hyperaldosteronism, hypocortisolism (Addison's disease), hyperthyroidism (Grave's disease), hypothyroidism, colorectal polyps, gastritis, gastric and duodenal ulcers, ulcerative colitis, and Crohn's disease; metabolic diseases such as obesity and a number of inflammatory diseases due to interleukin over-expression. For diagnostic and prognostic purposes, the expression of the protein of the invention could be investigated using any of the Northern blotting, RT-PCR or immunoblotting methods described herein and compared to the expression in control individuals. For prevention and/or treatment purposes, the protein of the invention may be overexpressed using any of the gene therapy methods known to those skilled in the art including those described herein. For example, expression of the protein of the invention can be upregulated by infecting tumor cells with a retroviral or an adenoviral vector which expresses the desired protein at higher levels necessary for suppression of mutation in the Rb gene or in other oncogenic or tumor suppressor genes.

Another related embodiment relates to the use of SEQ ID NO: 406, SEQ ID NO: 165, its complement, or any part thereof to develop antagonists of the protein of the invention. These antagonists could be antisense oligonucleotides, triple helices, ribozymes, small molecules or antibodies, especially neutralizing antibodies binding to the WD-repeats of the invention, and may be used to treat disease and conditions caused by abnormally low transcription. These conditions include accelerated aging syndromes such as Cochin's syndrome, Ataxia telangiectasia and Werner's syndrome as well as age-associated diseases as well as "early onset" forms of diseases associated with old age such as dementia and Parkinson's disease.

In another embodiment, the invention relates to methods and compositions using the protein of the invention or part thereof as a marker protein to selectively identify tissues, preferably colon and prostate. For example, the protein of the invention or part may be used to synthesize specific antibodies using any techniques known to those skilled in the art including those described therein. Such tissue-specific antibodies may then be used to identify tissues of unknown origin, for example, forensic samples, differentiated tumor tissue that has metastasized to foreign bodily sites, or to differentiate different tissue types in a tissue cross-section using immunochemistry.

Protein of SEQ ID NO: 414 (internal designation 188-27-3-0-G1-CS)

The 389 amino-acid long protein of SEQ ID NO: 414, expressed in brain, fetal brain, placenta and testis, over-expressed in brain and encoded by the cDNA of SEQ ID NO: 173 is homologous to SIRTUIN-2 (SIRT2) (Trembl accession number: Q9Y6E9) and Silent Information

Regulator 2-like protein (SIR2L) (Trembl accession number: Q9UNT0) that belong to the Silent Information Regulator type 2 (SIR2) family. In addition, the protein of the invention presents the Pfam signature for members of the SIR2 family (amino acids 84-268). Furthermore, the protein of the invention displays two characteristic motifs highly conserved among all members of the SIR 2 family that have been shown to be essential in the SIR2 silencing function (Moirá M. et al., Genetics, 154:1069-1083 (2000)). These motifs are from positions 84 to 98 and from positions 165 to 170 of the protein of SEQ ID NO: 414 and correspond to GAG(I/V)SxxxG(I/V)PDFERS and (Y/I)TQNID patterns respectively. The protein of SEQ ID NO: 414 also has conserved cysteines residues at positions 195, 200, 221 and 224, covering a domain thought to be either a DNA-binding zinc-finger motif (Prodom prediction PD002659, from positions 195 to 224) or an enzymatic domain (or an enzyme cofactor) (Moirá M. et al., Genetics, 154:1069-1083 (2000)).

The cDNA of SEQ ID NO: 173 encoding the protein of the invention differs from the one encoding the SIRT2 protein by a supplementary exon between positions 147 to 195. This exon modifies the initiation codon of the protein and extends the ORF in its N-terminal part by 16 amino acids. Moreover, amino residues in positions 20 and 21 of the protein of the invention (respectively an alanine and a glutamine residue) are substituted from a glutamine and a tyrosine residue (positions 4 and 5) of the SIRT2 protein. Thus, the protein of the invention is a new isoform of SIR2 resulting from alternative splicing. The protein of the invention of SEQ ID NO: 414 is also 37 amino acids longer than the SIRT2 protein at its N-terminal end.

Regulation of gene expression by alterations in chromatin structure is a universal mechanism in eukaryotic cells, responsible for maintaining patterns of gene expression throughout the development of multicellular organisms. Silencing has been studied most extensively in *S. cerevisiae* (yeast). Among the SIR genes, SIR2 is the most evolutionarily conserved, and a number of genes with homology to SIR2 have been identified (Frye R et al., Biochem. Biophys. Res. Commun., 273:793-798 (2000)). Presence of Homologues of SIR2 (*HSTs*) in organisms from bacteria to humans suggests that SIR2's silencing mechanism might be conserved. SIR2 was originally discovered to influence mating-type control in haploid cells by locus-specific transcriptional silencing. It has also been suggested that SIR2 and its homologs play additional roles in suppression of recombination, chromosomal stability, metabolic regulation, meiosis, and aging (for a review: see Gartenberg, Curr. Opin. Microbiol. 3:132-137 (2000)).

Proteins of the SIR2 family are also thought to be either enzymes or enzyme cofactors. First, Landry and collaborators have shown that members of SIR2 family catalyze histone deacetylation in a reaction that requires NAD, thereby distinguishing them from previously characterized deacetylases. This enzyme is active on histone substrates that have been acetylated by both chromatin assembly-linked and transcription related acetyltransferases (Landry et al., Proc. Natl. Acad. Sci. 97:5807-5811 (2000)). Discovery of an intrinsic deacetylation activity for the conserved SIR2 family provides a mechanism for modifying histones and other proteins to regulate

transcription and diverse biological process. Secondly, the study of a human SIR2 family member (hSirT2) was found to have a mono-ADP ribosylation activity *in vitro* (Frye R et al., Biochem. Biophys. Res. Commun., 260: 273-279 (1999)). Among potential substrates for mono-ADP ribosylation are histones and RNA Pol I, modification of which correlates with enhanced rDNA

5 transcription.

It is believed that the protein of SEQ ID NO: 414 or part thereof plays a role in gene silencing, suppression of recombination, chromosomal stability, metabolic regulation, meiosis, and aging, probably as a member of the SIR2 protein family. Particularly, the protein of the invention may deacetylate substrates, preferably acetylated histones and acetyltransferases, either directly or indirectly as enzyme cofactors. Particularly, the protein of the invention may have a ribosylation activity, preferably a mono-ADP ribosylation activity, preferably on histones and RNA Pol I substrates, either directly or indirectly as enzyme cofactors. Additionally, the protein of the invention may be a DNA binding protein. Preferred polypeptides of the invention are polypeptides comprising the amino acids of SEQ ID NO: 414 from positions 1 to 16, 84 to 98, 165 to 170, 195 to 224, and 84-268 as well as fragments of SEQ ID NO: 414 containing at least one cysteine residues located in positions 195, 200, 221 or 224 of SEQ ID NO: 414. Other preferred polypeptides of the invention are fragments of SEQ ID NO: 414 having any of the biological activities described herein. The deacetylation activity of the protein of the invention or part thereof may be assayed using any of the assays known to those skilled in the art including those described in Laundry et al., supra.

15 20 The ribosylation activity of the protein of the invention or part thereof may be assayed using any of the assays known to those skilled in the art including those described in Frye et al, (1999). The nucleic acid binding activity of the protein of the invention or part thereof may be assayed using any of the assays known to those skilled in the art including those described in US patent 6,013,453.

The invention relates to methods and compositions using the protein of the invention or part thereof to silence gene expression. In a preferred embodiment, the protein of the invention or part thereof or derivative thereof is added in an effective amount to an *in vitro* culture to inhibit gene expression and thus cell proliferation using molecular biology techniques known to those skilled in the art allowing the import of the protein from the extracellular medium to the cell's nucleus. In another embodiment, eukaryotic cells are genetically engineered in order to express the protein of the invention or part thereof under specific conditions in order to prevent further proliferation of such cells upon demand such as infection, transformation, end of a production process, differentiation, etc...

25 30

The invention relates to methods and compositions using the protein of the invention or part thereof to deacetylate substrates, alone or in combination with other substances. Such substrates are acetylated substrates, preferably acetylated histones and acetyltransferases. For example, the protein of the invention or part thereof is added to a sample containing the substrate(s) in conditions allowing deacetylation, and allowed to catalyze the deacetylation of the substrate(s). In a preferred

35

embodiment, the deacetylation is carried out using a standard assay such as those described in Laundry et al, supra. Deacetylated histones obtained by this method may be mixed with purified naked DNA (plasmid preparations for example) in order to reconstitute chromatine-like structures *in vitro*. Such structures are of great interest in the study of enzymatic factors involved in

5 transcription and replication.

Another embodiment of the present invention relates to composition and methods of using the protein of the invention or part thereof to develop assays for *in vitro* screening of inhibitors directed against the encoded deacetylase activity using any technique known to those skilled in the art including those described herein. Such deacetylase inhibitors are of great potential as new drugs
10 due to their ability to influence transcriptional regulation and to induce apoptosis or differentiation in cancer cells. Preferably, the protein of the invention, expressed in prokaryotic or eukaryotic systems according to methods known to those skilled in the art, may be mixed *in vitro* with a simple fluorescent substrate like an aminocoumarin derivative of an acetylated lysine, and different putative inhibitors. The coumarin derivative is then quantitated using a reverse-phase HPLC-system
15 with a fluorescence detector. Such an approach has been previously developed by Hoffmann and collaborators (Hoffmann et al., Nucl. Acids Res. 27:2057-2058 (1999); Hoffmann et al., Pharmazie 55:601-606 (2000)).

The invention relates to methods and compositions using the protein of the invention or part thereof to bind to nucleic acids, preferably DNA, alone or in combination with other substances.
20 For example, the protein of the invention or part thereof is added to a sample containing nucleic acid in conditions allowing binding, and allowed to bind to nucleic acids. In a preferred embodiment, the protein of the invention or part thereof may be used to purify nucleic acids such as restriction fragments. In another preferred embodiment, the protein of the invention or part thereof may be used to visualize nucleic acids when the polypeptide is linked to an appropriate fusion
25 partner, or is detected by probing with an antibody. Alternatively, the protein of the invention or part thereof may be bound to a chromatographic support, either alone or in combination with other DNA binding proteins, using techniques well known in the art, to form an affinity chromatography column. A sample containing nucleic acids to purify is run through the column. Immobilizing the protein of the invention or part thereof on a support advantageous is particularly for those
30 embodiments in which the method is to be practiced on a commercial scale. This immobilization facilitates the removal of the protein from the batch of product and subsequent reuse of the protein. Immobilization of the protein of the invention or part thereof can be accomplished, for example, by inserting a cellulose-binding domain in the protein. One of skill in the art will understand that other methods of immobilization could also be used and are described in the available literature.

35 Still another embodiment of the present invention relates to composition and methods of using the protein of the invention or part thereof to identify genes or regions of the human genome silenced by the protein of the invention or part thereof. Genomic DNA derived from patients with

pathologies such as cancer and metabolic disorders, or from elderly people may be compared to those extracted from respective controls for their ability to bind the protein of the invention. As described previously, the protein of SEQ ID NO: 414 displays a putative zinc finger domain susceptible to bind DNA sequences near regions or genes to silence. The protein of the invention or
5 part thereof may be bound to a chromatographic support, using techniques well known in the art, to form an affinity chromatography column. A sample containing a mixture of human genomic DNA digested by endorestriction enzymes is run through the column. After extensive washings the bound DNA is eluted and further subcloned in classical cloning vectors known to those skilled in the art. Immobilizing the protein of the invention or part thereof on a support is particularly advantageous
10 for those embodiments in which the method has to be practiced routinely. This immobilization facilitates the removal of DNAs from the batch of resin coupled protein after binding, and allows subsequent re-use of the protein. Immobilization of the protein of the invention or part thereof can be accomplished, for example, by inserting any matrix-binding domain in the protein according to methods known to those skilled in the art. The resulting fusion product including the protein of the
15 invention or part thereof is then covalently, or by any other means, bound to a protein, carbohydrate or matrix (such as gold, "Sephadex" particles, and polymeric surfaces).

Another embodiment of the invention relates to methods of preparing antibodies directed against the protein of the invention or part thereof. Such antibodies may be used in co-immunoprecipitation procedures that enrich for chromatin fragments containing binding sites for the
20 protein of the invention. This method may identify genes or regions of the human genome silenced by the deacetylase activity of the protein of the invention. Preferably, in samples containing fragments of native chromatin, antibodies directed against 414 and coupled to protein A or protein G sepharose beads are added to the mixture. Immunoprecipitation conditions are those known to those skilled in the art. After washings DNA fragments co-precipitated with 414 are extracted and
25 further subcloned in routinely used cloning vectors. These DNA fragments are either sequenced and/or used as probes to screen genomic libraries. This procedure is very similar to the one used by Gould and collaborators to enrich for embryonic chromatin fragments containing sites for the homeotic Ubx protein (Gould et al., Nature 348:308-312 (1990)).

A preferred embodiment of the invention relates to compositions or methods using SEQ ID
30 NO: 414, SEQ ID NO: 173 or part thereof to diagnose, treat and/or prevent develop disorders caused by the expression of "disease causing" genes. The number of pathologies and conditions that could be treated by the protein of the invention is potentially huge and unlimited. Favored disorders linked to dysregulation of gene transcription such as cancer and other disorders relating to abnormal cellular differentiation, proliferation, or degeneration, including hyperaldosteronism,
35 hypocortisolism (Addison's disease), hyperthyroidism (Grave's disease), hypothyroidism, colorectal polyps, gastritis, gastric and duodenal ulcers, ulcerative colitis, and Crohn's disease; viral infection especially HIV and viral hepatitis (i.e. expression of viral proteins), metabolic diseases such as

obesity and a number of inflammatory diseases due to interleukin over-expression. For diagnostic purposes, the expression of the protein of the invention could be investigated using any of the Northern blotting, RT-PCR or immunoblotting methods described herein and compared to the expression in control individuals. For prevention and/or treatment purposes, the protein of the invention may be overexpressed using any of the gene therapy methods known to those skilled in the art including those described herein. For example, switching off "disease" genes may be achieved by, for example, directly targeting the protein of the invention or part thereof to the genes (such as oncogenes in cancers) that are over-expressed in order to silence their expression. This could be achieved by making a "chimera" protein in which the putative zinc-binding domain is replaced by a sequence known to bind to, or near the over-expressed gene as explained elsewhere in the application. Fusion proteins containing both the deacetylase activity and the specific DNA binding domain are obtained by methods of molecular biology well known to those skilled in the art. The corresponding eukaryotic expression vectors may be used in gene therapy in the cases of cancer, metabolic disorders, aging and any disorder where a gene is over-expressed. Such recombinant cDNA may be introduced in the well known adenoviral vectors used in cancer therapy (for a recent review on the use of replicative adenoviruses for cancer therapy : Alemany et al., Nat. Biotechnol. 18:723-727 (2000)).

Another related embodiment relates to the use of SEQ ID NO: 414, SEQ ID NO: 173, its complement, or any part thereof to develop antagonists of the protein of the invention and of the SIR complex. These antagonists could be antisense oligonucleotides, triple helices, ribozymes, small molecules or antibodies and may be used to treat disease and conditions caused by abnormal gene silencing. These conditions include accelerated aging syndromes such as Cochin's syndrome, Ataxia telangiectasia and Werner's syndrome as well as age-associated diseases as well as "early onset" forms of diseases associated with old age such as dementia and Parkinson's disease.

In another embodiment, the invention relates to methods and compositions using the protein of the invention or part thereof as a marker protein to selectively identify tissues, preferably brain tissues. For example, the protein of the invention or part may be used to synthesize specific antibodies using any techniques known to those skilled in the art including those described therein. Such tissue-specific antibodies may then be used to identify tissues of unknown origin, for example, forensic samples, differentiated tumor tissue that has metastasized to foreign bodily sites, or to differentiate different tissue types in a tissue cross-section using immunochemistry.

Protein of SEQ ID NO:298 (182-1-2-0-D12-CS)

The protein of SEQ ID NO:298, encoded by the cDNA of SEQ ID NO:57, is homologous to proteins of the fibroblast growth factor family (FGF). Specifically the amino acid sequence of SEQ

ID NO:298 is identical to the recently described FGF-23. The protein of the invention is strongly expressed in the fetal liver.

The protein of the invention presents the pfam signature for fibroblast growth factors (positions 48 to 129). High resolution X-ray structures of crystals of both FGF-1 and FGF-2 have been reported and reveal a "beta trefoil" topology, comprising 12 strands linked to form a three-fold symmetrical structure made up of four-stranded antiparallel beta sheet (see Faham S. et al. - Curr Opin Struct Biol. - 1998, 8(5): p578-586). On the basis of sequence conservation, it seems very likely that all members of the FGF family have related 3-dimesional structures. Preferred polypeptides of the invention are those that comprise amino acids 39 to 45; 51 to 56; 60 to 64; 71 to 77; 82 to 87; 92 to 97; 101 to 105; 113 to 119; 124 to 130; 142 to 147; 151 to 155 and/or 167 to 172, which by homology with other members of the FGF family make up the 12 beta pleated sheets characteristic of the FGF family (White K. et al. - Nat Genet. - 2000; 26(3): pp. 345-348). Furthermore, as within these regions a number of amino acids from SEQ ID NO:298 are conserved in over 80% of human FGFs (after sequence alignment), the most preferred polypeptides of the invention comprise amino acids 42, 53, 63, 83, 85, 87, 93, 96, 101, 113, 115, 124, 127 and/or 129. Other preferred polypeptides of the invention are any fragment of SEQ ID NO:298 having any of the biological activities described herein.

Cytokines are a heterogeneous group of polypeptide mediators associated with numerous functions, including immune system and inflammatory responses. The cytokine families include, but are not limited to, Interleukins, Chemokines, Tumor necrosis factors, Interferons, Colony stimulating factors, Neurotrophins, neuropoietins and growth factors (of which the FGF family is a member). Fibroblast growth factors (FGFs) were first characterized, in the mid 1970s, as mitogens of cultured fibroblasts. Since then more then 20 different FGFs have been identified. Fibroblast growth factors belong to a family of proteins called growth factors (other members of this family include EGF, PDGF, TGFs and ECGF). The biological effects of FGFs are mediated by association with 3 biochemically distinct partners: heparan sulfate proteoglycans, a low affinity transmembrane FGF-binding protein and high-affinity transmembrane FGF receptors of the tyrosine-kinase class. Transfection and reconstruction experiments have shown that intracellular signal transduction is triggered by activation of FGF receptor kinase activity. Activation is brought about by receptor oligomerization, which is mediated by the association of heparan sulfate proteoglycans with the ligand (FGF) and of the ligand with the receptor itself (Faham S. et al. - Curr Opin Struct Biol. - 1998, 8(5): p578-586). Longer heparin-derived oligosaccharides generally exhibit tighter binding to FGF. The relationship between heparin length, biological activity and FGF binding has been extensively studied and there is general agreement that longer heparin oligosaccharides tend to be more biologically active. FGFs are members of a family with a broad range of biological activities involving cell growth and differentiation (including angiogenesis, morphogenesis and wound healing); cell survival, replication, adhesion and mobility. FGFs have been found to be potent

growth factors for a number of cell types including, but not limited to fibroblasts, endothelial cells, smooth muscle cells, keratinocytes, osteoblasts and neurons.

Clearly, FGF biology is potentially very complex, involving multiple ligands, receptors and cofactors, each expressed with different spatial and temporal patterns and distinct kinetics in the course of normal development. Considerable efforts have been expended on the creation of different types of animal models for the analysis of FGF function in vivo. These studies clearly indicate that FGF signaling is involved in a number of different processes at different stages of development and is critical in early developmental stages (FGF-4 and FGF receptor 1 homozygous null mutations both cause early lethality). FGF signaling has been found to be required for both branching morphogenesis of the lung and the establishment of the normal program of keratinocyte differentiation in the skin. FGF signaling has also been found to be involved in both the initial induction and sustained outgrowth of the limb bud during early limb development. Perhaps the most impressive illustration of this function of FGF signaling is the ability to induce supernumerary limb development in the chick by local application of FGF-soaked beads (Cohn M, et al. - Cell - 1995; 80: p739-746), thus indicating that a least some FGF-dependent processes are regulated by accessibility of an FGF ligand. FGFs are also capable of stimulating migration and differentiation of hepatic precursors.

Recently mutations in the FGF-23 gene were found to be associated with autosomal dominant hypophosphataemic rickets (ADHR), a genetically transmitted disease characterized by low serum phosphorus concentrations, rickets, osteomalacia, lower extremities deformation, short stature, bone pain and dental abscesses. It seems very likely that FGF signaling functions are involved in numerous aspects of morphogenesis, differentiation and other essential cellular mechanisms, and are thus likely involved in any of a large number of diseases and conditions associated with these processes.

Thus, it is believed that the protein of SEQ ID NO:298 is a member of the fibroblast growth factor family, and is thus involved in a large number of cellular and organismal processes, including, but not limited to, cell growth and differentiation, angiogenesis, morphogenesis, wound healing, cell survival, replication, adhesion and mobility.

One embodiment of the present invention relates to the use of the present polypeptides and polynucleotides to identify liver, heart, thyroid and parathyroid tissues, or cells derived from these tissues, since the protein of the invention is expressed therein (White K. et al. - Nat Genet. - 2000; 26(3): p345-348). Such detection of cells expressing the protein can be carried out in any of a number of ways, including the use of specific antibodies or antiserum generated against the protein using standard methods, as well as using polynucleotide probes specific for nucleic acids encoding the protein of the invention.

In another embodiment, the protein of the invention or part thereof can be used as a mitogen to stimulate the growth of a number of different cells types including, but not limited to, fibroblasts,

muscle cells, osteoblasts, keratinocytes and hepatocytes. The growth of cells can be stimulated in vitro, for example to promote the growth of cells cultured for the synthesis of recombinant proteins, or for ex vivo gene therapy applications. Another preferred application of this technique relates to the use of the protein of the invention or part thereof to generate in vitro tissues and organs including, but not limited to, skin, cartilage, and bone for transplants and grafts (Lancet – 1981, 1(8211):75-8)).

Another preferred embodiment of the invention relates to the use of the invention or part thereof to treat damaged tissues and organs. Members of the FGF family have been shown to induce the differentiation and growth of a number of different cell types. Thus the protein of the invention can be administered to treat pathologies and conditions that result from damage to cells, tissues or organs. These pathologies and conditions include but are not limited to bone fractures and bone defects Kimoto et al. – J Dent Res – 1998, 77(12): p1965-1969) (Solheim E – Int Orthop – 1998, 22(6), damage due to wounds (such as lesions of the skin and ulcers) (Debus E. – Zentralbl Chir – 2000, 125 (suppl 1): p49-55) (Szabo S. – Aliment Pharmacol Ther – 2000, 14(Suppl 1): p33-43), tissue damage due to ischemia (for example, in the brain and heart) (Simons M. – Circulation – 2000, 102(11): pE73-E86), cardiovascular diseases such as thrombosis and atherosclerosis (Bauters C. – Drugs – 1999, 58 (Spec No1): p11-15) and neurodegenerative diseases due to neuronal loss such as Parkinson's disease or Alzheimer's disease (Ebadi M – Neurochem Int – 1997, 30(4-5): p347-374) (Brundin P. – Cell Transplant – 2000, 9(2): p179-195).

In a most preferred embodiment, the polypeptides or polynucleotides of the invention can be used to diagnose, treat, or prevent disorders resulting from non-functional and/or mutated FGFs, such as Autosomal dominant hypophosphataemic rickets, which is associated with mutation of certain amino acids of FGF-23 (White K. et al. - Nat Genet. - 2000; 26(3): p345-348, which is hereby incorporated by reference in its entirety. Such disorders can be treated, for example, by administering a therapeutically effective amount of the protein of the invention or a polynucleotide sequence encoding the protein to a patient suffering from the disorder. Similarly, SEQ ID NO:298 or SEQ ID NO:57 or any part thereof can be used to develop diagnostic kits in order to diagnose, prevent and/or treat any other disease associated with FGF, for example pathologies associated with FGF overexpression.

In yet another embodiment, the protein of the invention or part thereof can be used to develop antagonists of FGF and/or FGF receptors in order to treat disorders associated with an over-activation of FGF pathways (for example, due to over production of FGF or overstimulation of FGF receptors). This is particularly true for pathologies such as cancers where some tumors secrete large quantities of FGF, such as prostate and breast cancers. Furthermore FGF antagonists can be useful in inhibiting tumor angiogenesis, which is an essential step in tumor growth. In the same way SEQ ID NO:57 or any part thereof could be used to generate antisense oligonucleotides. Antisense oligonucleotides block complementary mRNA, thus inhibiting the synthesis of the

protein encoded by the mRNA. These oligonucleotides can be used in in vivo or ex vivo treatment of the diseases caused or aggravated by overexpression of FGFs.

Protein of SEQ ID No: 396 (internal designation: 160-12-1-0-D10-CS)

The protein of SEQ ID No: 396 encoded by the cDNA of SEQ ID No: 155, overexpressed
 5 in brain and fetal brain, shows homology to members of the transmembrane 4 super family of proteins (TM4SF). The protein of the invention displays signatures characteristic of this family, namely the pfam domain from positions 66 to 273, the Prosite motif from positions 112 to 134 as well as the emotif domains from positions 108 to 127, 108 to 146, 129 to 150, 128 to 154, and 247 to 274. In addition, the protein of the invention has several predicted transmembrane domains: 103
 10 to 123, 130 to 150 and 245 to 265, with an additional predicted domain with lower certainty from positions 61 to 81. The protein of the invention has significant homology to a TM4SF member, the integral membrane CD81 antigen also known as TAPA1 (Target of Antiproliferative Antibody), except for its N-terminus. The transmembrane domains of the protein of the invention matches those described for CD81.

15 Members of the tetraspan family of proteins are associated with adhesion molecules and translate adhesive events into a regulation of cellular behaviour. TAPA-1 is a widely expressed protein found to influence adhesion, morphology, activation, proliferation and differentiation of B, T and other cells. TAPA-1 has two long hydrophilic domains of the molecule which are extracellular and located between four TM (Transmembrane region, TM1-4). The region between
 20 TM2 and TM3 is highly conserved in all tetraspanins. The protein is highly hydrophobic and contains a potential N-myristoylation site. TAPA1 functions by forming a complex on the cell surface and the antigenic epitope of the human TAPA1 is contained within a subregion of the second extracellular domain of the protein. Cell-surface expression of TAPA1 can be down-modulated by binding of antibodies (Levy 1991, J Biol Chem Aug 5;266(22):14597-602).

25 Mice lacking CD81 (not expressing TAPA1) have impaired antibody responses to protein antigens. This defect is specific to antigens that preferentially stimulate a T helper 2 response and is only seen with T cell-dependent antigens. Absence of CD81 on B cells is sufficient to cause the defect. Antigen-specific interleukin (IL) 4 production is greatly reduced in the spleen and lymph nodes of CD81-null mice compared with heterozygous littermates. The expression of CD81 on B
 30 cells is critical for inducing optimal IL-4 and antibody production during T helper 2 responses. CD81 is likely to have a greater role in the control of immune responses than in the development of immune cells (Maecker (1997) J Exp Med 1997 Apr 21;185(8):1505-10). CD81 on B cells has the capacity to promote IL-4 secretion from T cells. Costimulatory proteins such as B7-1 and B7-2 have been shown in some systems to have differential effects on cytokine secretion by T cells.
 35 CD81 on B cells, can control cytokine production by T cells. TAPA-1 has been implicated to play an important role in the regulation of lymphoma cell growth.

TAPA-1 is highly expressed in many neurons of the brainstem. TAPA is found in all glial cells, and the level of this protein correlates with their maturation (Sullivan et al., 1998, *J Comp Neurol* 1998 Jul 6;396(3):366-80). This protein is expressed by ependyma, choroid plexus, astrocytes, and oligodendrocytes. TAPA1 is dramatically upregulated during early postnatal
 5 development, at the time of glial birth and maturation. At embryonic day 18, the levels of TAPA are low, with most of the immunoreaction product being associated with the ependyma, choroid plexus, and the glia limitans. The amount of TAPA expressed in the brain increases with brain development, and at postnatal day 14 the protein levels approach those of the adult. This increase in the levels of TAPA at postnatal day 14 is due to upregulation in the gray matter and white matter.
 10 TAPA has been associated with reactive gliosis and the glial scar. The spatiotemporal expression pattern of CD81 by reactive microglia and astrocytes indicates that CD81 is involved in the glial response to spinal cord injury. It is suggested that the upregulation of TAPA is an integral component of glial scar formation (Peduzzi et al, *Exp Neurol*. 1999 Dec;160(2):460-8).

The levels of TAPA-1 are low in metastatic prostate tumors, expression of this protein in
 15 these cells appears to suppress metastatic behavior (Dong et al., 1995 *Science*. 12;268(5212):884-6.). Bivalent antibodies directed against these proteins can be used to enhance adhesion of different cell types: pre-B cells (Masellis-Smith and Shaw (1994) *J Immunol* 1994 Mar 15;152(6):2768-77), endothelial cells (Forsyth, 1991 *Immunology* Feb;72(2):292-6), and tumor cell mobility and invasiveness (Miyake et al., 1991 *J Exp Med*. Dec 1;174(6):1347-54.). In the nervous system, the
 20 migratory behavior of Schwann cells over biologically relevant substrates can be enhanced with the application of antibodies directed against certain TM (Anton et al (1995) *J Neurosci*. Jan;15(1 Pt 2):584-95). Antibodies directed against TAPA-1 depress the mitotic activity and induce an increase in cellular adhesion (Oren et al, 1990 *Mol Cell Biol* Aug;10(8):4007-15).

It is believed that the protein of SEQ ID NO: 396 or part thereof plays a role in cell
 25 adhesion, motility, metastasis, cell activation, signal transduction and the immune response, probably as a member of the TM4SF family. As a member of the tetraspanin family of proteins, the protein of SEQ ID No: 396 or part thereof is believed to mediate cellular interaction in lymphoid cells as well as non-hematolymphoid tissue to affect cell adhesion and migration, alter cell morphology and the activation state of a cell. Preferred polypeptides of the invention are
 30 polypeptides comprising the amino acids of SEQ ID NO: 396 from positions 66 to 273, 112 to 134, 108 to 127, 108 to 146, 129 to 150, 128 to 154, and 247 to 274. Other preferred polypeptides of the invention are fragments of SEQ ID NO: 396 having any of the biological activity described herein. The activity of the protein of the invention or part thereof may be assayed using any of the assays known to those skilled in the art including those describing a functional tissue assay used to define
 35 surface antigens regulating astrocyte growth (Eldon et al, 1996, *J Neurosci*, 16(17):5478); cellular function assays determining the involvement of the protein in signal transduction and cell adhesion

in the immune system (Levy et al, 1998 Ann. Rev. Immunol. 16:89-109, Virtaneva et al, 1994 Immunogenetics 39: 329-334).

An embodiment of the present invention relates to methods of using the protein of the invention or part thereof to identify and/or quantify membrane proteins, preferably integrins, lineage specific molecules, tetraspanins, and antibodies, in a biological sample, and thus used in assays and diagnostic kits for the quantification of such membrane proteins in tissue sample, and in mammalian cell cultures. The binding activity of the protein of the invention or part thereof may be assessed using the assay described in Shoshana et al, 1998, Annu. Rev. Immunol 16: 89-109; Maecker et al, 1998 PNAS 95: 2458-2462; Geisert et al, 1996, J of Neuroscience 16(17): 5478-5487 or any other method familiar to those skilled in the art. Preferably, a defined quantity of the protein of the invention or part thereof is added to the sample under conditions allowing the formation of a complex between the protein of the invention or part thereof and the membrane protein to be identified and/or quantified. Then, the presence of the complex and/or or the free protein of the invention or part thereof is assayed and eventually compared to a control using any of the techniques known by those skilled in the art.

In another embodiment, the invention relates to compositions and methods using the protein of the invention or part thereof to stimulate cell proliferation, preferably proliferation of lymphoid cells both in vitro and in vivo. For example, soluble forms of the protein of the invention or part thereof may be added to cell culture medium in an amount effective to stimulate cell proliferation.

In another embodiment, the invention relates to compositions and methods using the protein of the invention or part thereof or derivative thereof to stimulate antibody production either in vitro or in vivo. In a preferred embodiment, the protein of the invention or part thereof or derivative thereof may be added in an effective amount to stimulate antibody production to an in vitro culture of antibody-producing cells, such as hybridomas. In another preferred embodiment, the protein of the invention or part thereof or derivative thereof may be injected into an animal in order to increase the animal's antibody production to a protein of interest in the case of production of polyclonal antibodies.

In still another embodiment, the invention relates to compositions and methods using the protein of the invention or part thereof or derivative thereof to decrease cell adhesion either in vitro or in vivo. In a preferred embodiment, the protein of the invention or part thereof or derivative thereof may be added in an effective amount to prevent and/or inhibit cell adhesion to an in vitro culture of adherent cells in order to recover those adherent cells.

In still another embodiment, , the invention relates to compositions and methods using the protein of the invention or part thereof to treat and/or prevent cell-proliferative disorders, such as cancers,, via the prevention of metastatispreferably brain cancer, and disorders characterized by depressed immune response such as autoimmune diseases AIDS, allergy, type1 diabetes, systemic lupus erythematosus, chronic rheumatoid arthritis, juvenile rheumatoid arthritis, Sjogren's

syndrome, systemic scleriosis, mixed connective tissue disease and dermatomyositis, Hashimoto's disease, primary myxedema, thyrotoxia, pernicious anemia, ulcerative colitis, autoimmune atrophic gastritis, idiopathic Addison's disease, male infertility, Goodpasture's syndrome, acute progressive glomerular nephritis, myasthenia gravis, multiple myositis, pemphigus vulgaris, bullous

- 5 pemphigoid, sympathetic ophthalmia, multiple sclerosis, autoimmune hemolytic anemia, idiopathic thrombocytopenic purpura, postmyocardial infarction syndrome, rheumatic fever, lupoid hepatitis, primary biliary cirrhosis, Behcet's syndrome and Crest's syndrome, via the stimulation of antibody production and IL-4 secretion.

In another embodiment, the invention relates to methods and compositions using the protein
 10 of the invention or part thereof as a marker protein to selectively identify tissues, preferably from brain and fetal brain origin. For example, the protein of the invention or part may be used to synthesize specific antibodies using any techniques known to those skilled in the art including those described therein. Such tissue-specific antibodies may then be used to identify tissues of unknown origin, for example, forensic samples, differentiated tumor tissue that has metastasized to foreign
 15 bodily sites, or to differentiate different tissue types in a tissue cross-section using immunochemistry or any other technique known to those skilled in the art.

Protein of SEQ ID No: 296 (internal designation 181-3-3-0-B8-CS)

The protein of SEQ ID NO: 296 encoded by the cDNA of SEQ ID NO:55, overexpressed in fetal liver, is homologous to the whole domain IV-4 and IV-5 of the basement membrane-specific
 20 heparan sulfate proteoglycan core protein (perlecan), well conserved among *C.elegans*, mice and human (accession numbers Q06561, Q05793 and P98160 respectively). The 247-amino-acid-long protein of the invention, displays two putative hydrophobic stretches from positions 44 to 64 and 219 to 239 and a putative immunoglobulin domain from positions 141 to 197, homolog to the Ig domain 4 of the Ig repeat structure of domain IV of perlecan proteins. The protein of the invention
 25 displays also a putative secreted signal peptide from positions 6 to 21.

Basement membranes are specialized regions of extracellular matrix (ECM) containing a large number of different components, including laminin, collagen, nidogen and heparan sulfate proteoglycans (for a review see Bernfield et al., *Annu. Rev. Biochem.* 68:729-777 (1999)). Perlecan, a major basement membrane, plays important roles in many fundamental development
 30 and regenerative processes, including cell cohesion, adhesion and migration, signal transduction, and even gene regulation (Martin and Timpl, *Annu. Rev. Cell Biol.* 3:57-85 (1987)). The cDNA sequence of perlecan encodes a large core protein consisting of five structural domains, referred from I to V, with distinct motifs such as SEA modules (domain I), LDL class A modules (domain II), cysteine-rich LE modules (domain III), LAMB modules (domain V). Domain IV consists of Ig-
 35 like repeats (14 in mice, 21 in human perlecan) similar to those of neural cell adhesion molecules (N-CAMs). Glycosaminoglycan chains are mostly linked to domain I and have been shown to

participate with the core protein in its differential expression in tissues and development stages (Perrimon and Bernfield, Nature 404:725-728 (2000)).

The N-terminal fragment IV containing Ig modules from 1 to 8 show high-affinity binding to the two known nidogen isoforms, laminin1-nidogen1 complex (LN) and binding to heparin at physiological ionic strength (Hopf et al., Eur. J. Biochem. 259:917-925(1999)). An alteration study of the C. elegans perlecan show in vivo that mutations in Ig-like modules 3 and 4 induce a lethal phenotype by inhibiting the spatial distribution of the splice variants. Mutations inducing deletions in other Ig-like modules of perlecan domain IV does neither affect the isoform expression nor the spatial distribution, suggesting a crucial role of Ig-like modules 3 and 4 in muscle assembly and development stages (Mullen et al., Mol. Biol Cell 10:3205-3221 (1999)).

Several studies have shown the large presence of distinct perlecan isoforms through regulated alternative splicing in C. elegans (Rogalski et al, Genes Dev. 7:1471-1484 (1993), Rogalski et al, Genetics 139:159-169(1995)). Although splice variants have not yet been shown in human, Ig-like modules are encoded by multiple exons compatible with different combinatorial possibilities of expression (Cohen et al., P.N.A.S. 90:10404-10408 (1993)). Alternative splicing within Domain IV is associated with temporal and spatial differences in isoform expression. A subset of C.elegans isoforms are associated with body-wall muscles during embryogenesis and are required for nematode myofilament lattice assembly, which is very similar to assembly of focal adhesions in mammalian cell culture (Moerman and Fire, CSH labo. Press (1997)).

Basement membrane-like structure containing perlecan, collagen IV, laminin also plays a major role during liver differentiation by interacting with immature hepatocytes (Am. J. Path. 142:199-208 (1993)).

Perlecan have been implicated in a number of processes and diseases resulting from the alteration of its structure including glomerular filtration deficiencies such as proteinuria, diabetic glomerulopathies, nephrotic syndromes, Denys-Drash syndromes (Groffen et al., Nephrol. Dial. Transplant 14:2119-2129 (1999)), mitogenesis and angiogenesis diseases (Aviezer et al., Cell 79:1005-1013 (1994)), inflammation and tissue repair, ocular and skeletal defect syndromes, microbial pathogenesis through invasion. Perlecan core protein has binding epitopes for the basement membrane proteins nidogen-1, nidogen-2, and fibulin-2, as well as for Alzheimer's beta-amyloid protein (Snow et al., Arch. Biochem. Biophys. 320:84-95 (1995)) and platelet growth factor.

It is believed that protein of SEQ ID NO: 296 or part thereof is a membrane basement-like protein, preferably a human isoform of the perlecan protein. Thus, the protein of the invention plays an important role in membrane integrity and interactions with other basement proteins and particularly with nidogen-1 and 2, LN complexes and heparin compounds. Besides, the protein of the invention is thought to participate in the interactions with cellular receptors such as integrins, with cytokine release and proteolysis, with regulation of angiogenesis, wound healing and tumor

invasion. Being overexpressed in the fetal liver, the protein of the invention is thought to participate in the differentiation, migration and adhesion of hepatocytes through its spatial and temporal expression during embryogenesis. Preferred polypeptides of the invention are polypeptides comprising the amino acids of SEQ ID NO: 296 from positions 141 to 197. Other preferred
5 polypeptides of the invention are fragments of SEQ ID NO: 296 having any of the biological activity described herein. The activity of the protein of the invention or part thereof may be assayed using any of the assays known to those skilled in the art including those described in Hopf et al., Euro. J. Biochem. 259:917-925(1999) for binding assays with other membrane proteins and in Rescan et al., Am. J. Path. 142:199-208 (1993) for protein assays (Immunocytochemistry and ELISA).

10 In one embodiment, the invention relates to methods and compositions using the protein of the invention or part thereof as a new marker protein to selectively identify embryogenic stages, preferably in liver tissues. For example, the protein of the invention or part thereof may be detected using specific antibodies able to bind to the protein using any technique known to those skilled in the art. Such tissue-specific antibodies may then be used to identify embryogenic cells with
15 dysregulated membrane components such as in differentiated tumor cells or to differentiate different cell types in a tissue cross-section using immunocytochemistry. For example, the amount of the protein of the invention in embryogenic cells reflecting the characterized overexpression activity is measured and compared to that of a normal cell using a specific antibody detected by fluorescence (FACS, confocal microscopy,...) or any other detection methods skilled in the art.

20 In another embodiment the invention relates to methods and compositions using the protein of the invention or part thereof for the diagnosis of a disorder associated with overexpression of the protein of the invention, preferably but not limited to perlecan associated tumors such as human melanoma, proliferative diseases, glomerular filtration deficiencies such as proteinuria, diabetic glomerulopathies, nephrotic syndromes, Denys-Drash syndromes, mitogenesis and angiogenesis
25 diseases, inflammation and tissue repair, ocular and skeletal defect syndromes, microbial pathogenesis through invasion. The expression of the protein of the invention could be investigated using any methods well known to those skilled in the art, including Northern blotting, RT-PCR or immunoblotting using specific antibodies binding to the protein of the invention.

In still another embodiment the protein of the invention or part thereof could be used as a
30 mitogen to stimulate the growth and differentiation of a number of different cell types including but not limited to fibroblasts, muscle cells, osteoblasts, keratinocytes and hepatocytes. In a preferred embodiment, the protein of the invention or part thereof is used in in vitro cultures such as those used for synthesis of recombinant proteins. The protein of the invention or part thereof is added to the culture in an amount effective to stimulate proliferation and/or differentiation. A more
35 preferred application of this technique relates to the use of the protein of the invention or part thereof in generating or repairing in vitro tissues and organs such as but not limited to skin, cartilage,

and bone for transplants and grafts (Lancet – 1981, 1(8211):75-8), which disclosure is hereby incorporated by reference in its entirety).

In another embodiment, an antagonist of the protein of SEQ ID NO:296 may be administered to a subject to treat or prevent a cell proliferative disorder. Such disorders may include, but are not limited to, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. In one aspect, an antibody which specifically binds to the protein of the invention may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express the protein of the invention. In another example, antisense nucleotides, triple helices, Genetic Suppressor Elements (GSE), ribozymes designed from nucleotides encoding the protein of the invention or part thereof using any methods to those skilled in the art are administered to inhibit the expression of the protein of the invention.

Protein of SEQ ID NO: 410 (internal designation 179-9-4-0-B8-CS)

The protein of SEQ ID NO: 410 encoded by cDNA of SEQ ID NO: 169 found in fetal kidney is homologous to the proteins of ankyrin family protein and the proteins containing a characteristic ankyrin repeated motif (pfam accession number : PF00023). The protein of the invention shows homology with human ankyrin proteins (PIR accession number A35049 ; SP-TREMBL accession number : Q99407), ankyrins from several different eucaryote species (*Drosophila melanogaster* : STR accession number Q9VAU5 ; mouse : STR accession number Q61302 and SwissProt accessiion number Q02357 ; cow : STR accession number AAF61702 ; *Arabidopsis thaliana* : STR accession number Q9ZQ79) and ankyrins from procaryote species (*Paramecium bursaria* *Chlorella* virus : STR accession number STR Q41164).

In addition, the protein of the invention shows homology with other proteins containing ankyrin repeat motif. The ankyrin repeat motif is a 33 amino acid motif and has an L-shaped structure consisting of two alpha helices following the beta hairpin loop (Gorina et al., Science.274-1005 (1996)). Examples of proteins comprising ankyrin repeats include: channels , enzymes toxins , transcription factors (Palek et al., Semin. Hematol. 27:290-332 (1990)), tankyrase (Smith et al., Science.282:1484-1487 (1998)) , multiple proteins involved in signal transduction, in particular integrin-linked kinases (Huang et al., Int. Mol. Med. 3:563-572 (1999)), inhibitors of cyclin-dependent kinases (Baumgartner et al., Structure. 6:1279-1290 (1998)), death-associated protein

kinase involved in apoptosis (Raveh et al., *Proc. Natl. Acad. USA.* 15:1572-1577 (2000)) and many others.

The ankyrin motif is also found in the protein of the invention (position 47 to 79).

Ankyrins are peripheral membrane proteins which have been found in erythrocyte, kidney
5 and neuronal cells of mammals. Cells contain a cytoskeleton that links intracellular compartments with each other and the plasma membrane. Associations between the cytoskeleton and the lipid membranes bounding these compartments involve spectrin, ankyrin, and integral membrane proteins. Spectrin is a major component of the cytoskeleton and acts as a scaffolding protein. Similarly, ankyrin acts to tether the actin-spectrin moiety to membranes and to regulate the
10 interaction between the cytoskeleton and membranous compartments. Different ankyrin isoforms are specific to different organelles and provide specificity for this interaction. Genes coding for three different mammalian ankyrins (ankyrin_R, ankyrin_B and ankyrin_C) have been cloned. Ankyrin_R was originally identified as part of the erythrocyte membrane skeleton, and was recently also localized to the plasma membrane of a subpopulation of post mitotic neurons in rat brain (Lambert,
15 et al., 1993, *J. Neurosci.*, 13, 3725-3735). Ankyrin_B is a developmentally regulated human brain protein which has two alternatively spliced isoforms with molecular masses of 220 kilodaltons (kD) and 440 kD (Kunimoto, et al., 1991, *J. Cell Biology*, 115, 1319-1331). Ankyrin_C is a more recently isolated human gene that encodes two neural-specific ankyrin variants (480 kD and 270 kD), which have been localized to the axonal initial segment and node of Ranvier (Kordeli, et al., 1995, *J. Biol.*
20 *Chem.*, 270, 2352-2359). Studies on mammalian ankyrins indicate that ankyrins bind a variety of proteins which have functions involved with the anion exchanger (Drenckhahn, et al., 1988, *Science*, 230, 1287-1289), Na⁺/K⁺-ATPase, amiloride-sensitive sodium channel in kidney (Smith, et al., 1991, *Proc. Natl. Acad. Sci. U.S.A.*, 88, 6971-6975), voltage dependent sodium channel of the brain and the neuromuscular junction (Srinivasan, et al., 1988, *Nature*, 333, 177-180), and
25 nervous system cell adhesion molecules (Davis, et al., 1994, *J. Biol. Chem.*, 269, 27163-27166).

Analyses of mammalian ankyrins have revealed that these large proteins are divided into three functional domains. These include an N-terminal membrane-binding domain of about 89-95 kD, a spectrin-binding domain of about 62 kD, and a C-terminal regulatory domain of about 50-55 kD. The membrane-binding domain is primarily comprised of tandem repeats of about 33 amino
30 acids each. This domain usually has about 22-24 copies of these repeats. The repeat units appear to function in binding to membrane proteins such as anion exchangers, sodium channels, and certain adhesion molecules. The spectrin-binding domain, as the name implies, functions in binding to the spectrin-based cytoskeleton of cells positioned inside the plasma membrane. Finally, the regulatory domain, which is the most variant domain among the different ankyrins that have been studied,
35 appears to function in as a repressor and/or an activator of the protein-binding activities of the other two domains. Some of the variability seen in this domain among different ankyrin species appears to be the result of alternative splicing of nascent transcripts. The regulatory domain can respond to

cellular signals, allowing remodeling of the cytoskeleton during the cell cycle and differentiation (Lambert, S. and Bennett, V. (1993) Eur. J. Biochem. 211:1-6). Ankyrin may be target for action of parasites. Erythrocyte ankyrin is cleaved by parasite proteases of plasmodium falciparum destabilizing erythrocyte membrane skeleton which facilitates parasite release (Raphael et al., Mol
5 Biochem Parasitol. 110(2):259-272 (2000)). Recently, novel ankyrin proteins have been isolated from Dirofilaria and Brugia which may be useful in protecting animals, including humans, from diseases caused by parasitic helminths (United States Patent No. 6,063,599).

Ankyrin sequences have been identified in various libraries, at least 50% of which are associated with cancer and at least 23% of which are associated with the immune response. Of
10 particular note is the expression of ANFP in reproductive and hematopoietic/immune, and gastrointestinal tissues. See United States Patent 5,989,863.

It is believed that the protein of SEQ. ID. NO: 410 is a member of the family of human ankyrin proteins and as such plays a role in regulating the interaction between the cytoskeleton and membranous components. The identification of a new member of the ankyrin family and the
15 polynucleotides encoding it addresses a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of autoimmune/inflammatory, cell proliferative, vesicle trafficking disorders and in modulating the response to infectious diseases.

Preferred polypeptides of the invention are polypeptides comprising the amino acids from positions 47 to 79. Other preferred polypeptides are fragments of SEQ.ID.NO: 410 having the
20 desired biological activity. Further included in the invention are the polypeptides encoded by the human cDNA of clone 179-9-4-0-B8-CS. The polypeptides of SEQ ID NO: 410 may be interchanged with the corresponding polypeptides encoded by the human cDNA of clone 179-9-4-0-B8-CS. Further included in the invention are polynucleotides encoding said polypeptides. Preferred polynucleotides are those of SEQ ID NO: 169 and of the human cDNA of clone 179-9-4-
25 0-B8-CS.

The invention also encompasses variants of the protein of the invention. A preferred variant is one which has at least about 80%, more preferably at least about 90%, and most preferably at least about 95% amino acid sequence identity to the amino acid sequence of SEQ.ID.NO: 410, and which contains at least one functional or structural characteristic of ankyrin.

30 In a particular embodiment, the invention encompasses a polynucleotide sequence comprising the sequence of SEQ ID NO: 410, as well as variants of that sequence. Variants which encode at least one functional region characteristic of the ankyrin protein of the present invention are encompassed. Codon usage may be varied according to standard techniques in order to enhance expression in various hosts.

35 In one embodiment, the protein of SEQ.ID.NO: 410 or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of ankyrin. Examples of such disorders include, but are not limited to,

autoimmune/inflammatory disorders such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis. autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED),

5 bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis,

10 pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjogren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; cell proliferative disorders such as actinic keratosis,

15 arteriosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver,

20 lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; and vesicle trafficking disorders such as cystic fibrosis, glucose-galactose malabsorption syndrome, hypercholesterolemia, diabetes mellitus, diabetes insipidus, hyper- and hypoglycemia, Grave's disease, goiter, and Cushing's disease, ulcerative colitis, and gastric and duodenal ulcers.

25 In another embodiment, a vector capable of expressing the protein of SEQ.ID.NO: 410 or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of ankyrin including, but not limited to, those described above.

In a further embodiment, a pharmaceutical composition comprising a substantially purified

30 protein of SEQ.ID. NO. 410 or a portion of the protein in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of the same or a similar protein including, but not limited to, those provided above.

In still another embodiment, an agonist of the protein of SEQ. ID. NO. 410 which

35 modulates the activity of the protein may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of the protein, or other ankyrin proteins, including, but not limited to, those listed above.

In another embodiment, the polypeptide of SEQ. ID. NO. 410 may be used to produce antagonists using methods which are generally known in the art. In particular, purified polypeptide may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind ankyrin proteins. Neutralizing antibodies (i.e., those which inhibit dimer
5 formation) can also be prepared for therapeutic use.

In a further embodiment, an antagonist of the protein of SEQ. ID. NO. 410 may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of the same protein or other members of the ankyrin family of proteins. Such disorders may include, but are not limited to, those discussed above. In one aspect, an antibody which specifically
10 binds the claimed polypeptide may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express the polypeptide.

In an additional embodiment, a vector expressing the complement of the polynucleotide of SEQ. ID. NO. 169 may be administered to a subject to treat or prevent a disorder associated with
15 increased expression or activity of ankyrin proteins including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. The combination of therapeutic agents may act synergistically to effect the
20 treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

In another embodiment of the invention, the polynucleotides encoding the polypeptide of SEQ. ID. NO. 410, or any fragment or complement thereof, may be used for therapeutic purposes.
25 In one aspect, the complement of the polynucleotide encoding the above-identified polypeptide may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be transformed with sequences complementary to polynucleotides encoding the polypeptide. Thus, complementary molecules or fragments may be used to modulate the activity of the claimed polypeptide or related ankyrin proteins, or to achieve regulation of gene function.

30 In another embodiment of the invention, the nucleotide sequence encoding the polypeptide of SEQ. ID. NO. 410 can be used to turn off the genes expressing the polynucleotide or related ankyrin proteins. In particular, a cell or tissue can be transformed with expression vectors which express high levels of the polynucleotide, or fragment thereof. Such constructs may be used to introduce untranslatable sense or antisense sequences into the cell. Expression vectors derived from
35 retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of the nucleotide sequences to a targeted organ, tissue, or cell population.

An additional embodiment of the invention relates to the administration of a pharmaceutical or sterile composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed above. The composition may be delivered via a variety of different routes.

5 In another embodiment, antibodies which bind the polypeptide of SEQ. ID. NO. 410 may be used for the diagnosis of disorders characterized by expression of ANFP, or in assays to monitor patients being treated with the polypeptide, other ankyrin proteins or agonists, antagonists, or inhibitors of the same. A variety of assay types, including ELISAs, RIAs, and FACS, can be used.

In another embodiment of the invention, the polynucleotide of SEQ. ID. NO. 169 itself,
10 may be used for diagnostic purposes. The polynucleotide can be used to generate oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs which are useful in diagnosis. The polynucleotide and related molecules may be used to detect and quantitate gene expression in biopsied tissues in which expression of the polypeptide of SEQ. ID. NO. 410 or other ankyrin polypeptides may be correlated with disease. The diagnostic assay may be used to determine
15 absence, presence, and excess expression of the polypeptides, and to monitor regulation of polypeptide levels during therapeutic intervention. Examples of diagnostic methods include Southern or Northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered ANFP expression. Such qualitative or quantitative
20 methods are well known in the art. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as targets in a microarray. The microarray
25 can be used to monitor the expression level of large numbers of genes simultaneously and to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, and to develop and monitor the activities of therapeutic agents.

In another aspect of the invention, the polypeptide may be used to stimulate the expression
30 of genes that have a role in organ and organ system development. Thus, in a preferred embodiment, the protein of the invention, a fragment, or derivative thereof, may be administered to a subject to treat or prevent developmental disorders.

In a further embodiment, the protein of the invention may be administered to a subject to treat or prevent a cardiovascular disorder. Such disorders can include, but are not limited to,
35 arteriosclerosis including atherosclerosis and nonatheromatous arteriosclerosis, hypertension, stroke, coronary artery disease, ischemia, myocardial infarction, angina pectoris, cardiac arrhythmias, sinoatrial node blocks, atrioventricular node blocks, chronic hemodynamic overload, aneurysm,

Jervell and Lange-Nielsen syndrome, and long QT syndrome. The protein of the invention may also be used as a marker of cardiac hypertrophy so it may be included in diagnosis kit for this disease.

In another embodiment of the invention, the polypeptide and/or polynucleotide may be used to inhibit cellular proliferation and to treat and/or diagnose disorders associated with cellular proliferation including but not limited to cancer.

In a further embodiment of the invention, an antagonist of the protein of the invention may be administered to a subject to treat or prevent a cancer. In one aspect, an antibody which specifically binds the protein of the invention may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express the protein of the invention.

In yet another embodiment, an antagonist of the protein of the invention may be administered to a subject to treat or prevent a neuronal disorder. Such a disorder may include, but is not limited to, akathisia, Alzheimer's disease, amnesia, amyotrophic lateral sclerosis, bipolar disorder, catatonia, cerebral neoplasms, dementia, depression, diabetic neuropathy, Down's syndrome, tardive dyskinesia, dystonias, epilepsy, Huntington's disease, peripheral neuropathy, multiple sclerosis, neurofibromatosis, Parkinson's disease, paranoid psychoses, postherpetic neuralgia, schizophrenia, and Tourette's disorder. In one aspect, an antibody which specifically binds the protein of the invention may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express the protein of the invention.

In another embodiment, the protein of the invention can be administered to a subject to treat or prevent a malaria. The protein of the invention may be used also in diagnosis of malaria.

In yet another embodiment, the polynucleotide and/or the polypeptide of the present invention can be used as a therapeutic composition capable of protecting an animal from a disease caused by a parasitic helminth. The polypeptide can be used a target for antiparasitic vaccines and drugs.

Ankyrin has been shown to underlie membrane proteins including CD44, the voltage-dependent sodium channel, Na^+/K^+ ATPase and the anion exchanger protein. It is believed that the formation of a direct connection between ankyrin and functionally important transmembrane proteins/membrane skeleton may be one of the earliest events to occur during signal transduction and cell activation. Thus, in a further embodiment, the polypeptide of the present invention can be used to disrupt the connection between ankyrin and the membrane thus affecting fundamental processes within the cell.

The polypeptide of the present invention can be further used to screen for compounds that inhibit or enhance the binding of ankyrin binding proteins and to affect the association between ankyrin and proteins, such as Alpha-Na,K-ATPase , which are critical to intracellular transport of ions and nutrients.

In yet another embodiment the regulatory domain of the polypeptide of SEQ. ID.NO. 410 or antagonists thereof can be used to enhance or disrupt the protein binding activities of the other domains.

Proteins of SEQ ID NOs: 385 and 416 (internal designations 105-021-3-0-C3-CS and 188-31-1-0-

5 E6-CS respectively)

The 354 amino acids protein of SEQ ID NO: 385 encoded by the cDNA of SEQ ID NO: 144 found in brain displays 6 kelch motifs (pfam accession number PF01344) at positions 20-66, 68-114, 116-162, 164-209, 211-265 and 270-316. Moreover, 4 residues conserved in over 90% of kelch family sequences are found in the protein of invention: di-glycine at positions 133-134, 10 tyrosine 148 and tryptophan 155. In addition, six residues separate the tyrosine 148 and the tryptophan 155: this feature is conserved in over 70% of kelch proteins (Adams et al., trends in cell biology, 10:17-24, 2000). The proteins of the invention encoded by the cDNA of SEQ ID NO: 144 is a polymorphic variant of the protein of SEQ ID NO: 416 encoded by the cDNA of SEQ ID NO: 175, thought to have the same functions and utilities.

15 *Drosophila* kelch is located in the ring canals which are actin-rich bridges. Kelch localizes to the rim of preformed canals and serves to maintain actin organization (Xue et al., Cell (72)681-693, 1993; Robinson et al., J. Cell Biol. (138)799-810, 1997). In mammalian sperm, calicin is located within an actin-negative structure termed the calyx, which is involved in the morphogenesis of the spermatocyte (von Bulow et al., Exp. Cell Res. (219)407-413, 1995). Calcin and a well- 20 structured calyx are both lacking in certain teratozoospermias, possibly indicating a central role for calicin in the organization of this structure (Courtot et al., Mol. Reprod. Dev. (28)272-279, 1991). In *Schizosaccharomyces pombe*, Ral2p acts down-stream of Ras1p in pathways that affect cell morphology, conjugation and sporulation. The spherical morphology and mating defects of ral2-null cells are complemented by overexpression of Ras1p, indicating a close functional interaction 25 between the two proteins (Fukui et al., Mol. Cell. Biol. (9)5617-5622, 1989). The transcription factor Nrf2 is sequestered by the kelch-repeat containing Keap1 protein under normal cellular conditions. The stimulation by agents such as diethylmaleate induces the translocation of Nrf2 to the nucleus to initiate the cytoprotective electrophilic counterattack response (Itoh et al., Genes Dev. (13)76-86, 1999). Lytic infection of cells by herpes simplex virus is initiated by binding of 30 virally encoded VP16 to HCF-1, a protein thought to have a normal role in cell-cycle progression. The HCF-VP16 complex then assembles with Oct-1 transcription factor on cis-regulatory targets in the HSV genome to initiate virus replication. (Wilson et al., Mol. Cell. Biol (17)6139-6146, 1997; Hughes et al., J. Biol. Chem. (274)16437-16443, 1999). Two recently discovered mammalian kelch-repeat proteins have extracellular roles. Human attractin appears to participate in normal immune 35 defence as a serum glycoprotein released by activated T cells. In coculture assays, attractin stimulates adhesion and spreading of monocytes, facilitating the development of T-cell clusters and

cellular immune responses (Duke-Cohan et al., Proc. Natl. Acad. Sci. U. S. A. (95)11336-11341, 1998). Attractin is orthologous to the extracellular domain of mouse mahogany, a large, multidomain, transmembrane protein that has been implicated in the homeostasis of energy metabolism by its suppressive effects on certain types of obesity in mice (Gunn et al., Nature 5 (398)152-156, 1999; Nagle et al., Nature (398)148-152, 1999).

Evidence for the importance of kelch repeat beta-propellers in protein function has also come from studies of natural and engineered loss-of-function mutations. *Caenorhabditis elegans* Spe-26 mutant spermatocytes fail to complete meiosis, contain multiple nuclei and show gross disorganization of actin filaments and organelles. For five out of the six alleles that have been 10 examined in detail, the mutations map within the kelch repeats (Varkey et al., Genes Dev. (9)1074-1086, 1995). Of particular interest are the point mutations in RAG-2 that have been identified in some cases of human B-cell-negative severe combined immuno-deficiency or of Omenn syndrome (Schwarz et al., Science (274)97-99, 1996; Villa et al., Cell (93)885-896, 1998). Very recently, the gigaxonin, a new member of the cytoskeletal BTB/kelch repeat family, is described as mutated in 15 giant axonal neuropathy (Bomont et al., Nat. Genet. (26) 370-374, 2000).

It is believed that the proteins of the invention are members of the kelch superfamily and, such as, play a role in the association with the actin cytoskeleton, the organization of cytoskeletal, plasma membrane or organelle structures, the coordination of morphology and growth, the gene expression, the viral pathogenesis, the immune defence. In particular, the proteins of invention are 20 highly expressed in brain and is believed to be related to the CNS disorders. Preferred polypeptides of the invention are polypeptides comprising the amino acids of the proteins of invention at positions 20-66, 68-114, 116-162, 164-209, 211-265 and 270-316. In one embodiment, the proteins of the invention or part thereof are used to modulate actin organization in cells thus affecting the cytoskeleton and cell function in general.

25 The invention also features compounds, e.g., proteins, which interact with the protein of the invention. Any method suitable for detecting protein-protein interactions may be employed for identifying transmembrane proteins, intracellular, or extracellular proteins that interact with the protein. Among the traditional methods which may be employed are co-immunoprecipitation, crosslinking and co-purification through gradients or chromatographic columns of cell lysates, or 30 proteins obtained from cell lysates, and the use of the proteins of the invention to identify proteins in the lysate that interact with it. For these assays, the protein of the invention can be full length or some other suitable protein polypeptide fragment. Once isolated, such an interacting protein can be identified and cloned and then used, in conjunction with standard techniques, to identify proteins with which it interacts. For example, at least a portion of the amino acid sequence of a protein 35 which interacts with the protein of the invention can be ascertained using techniques well known to those of skill in the art, such as via the Edman degradation technique. The amino acid sequence obtained may be used as a guide for the generation of oligonucleotide mixtures that can be used to

screen for gene sequences encoding the interacting protein. Screening may be accomplished, for example, by standard hybridization or PCR techniques. Techniques for the generation of oligonucleotide mixtures and the screening are well-known. ("PCR Protocols: A Guide to Methods and Applications," Innis et al., eds. Academic Press, Inc., NY, 1990).

- 5 Additionally, methods can be employed which result directly in the identification of genes which encode proteins that interact with the protein of the invention. These methods include, for example, screening expression libraries, in a manner similar to the well known technique of antibody probing of lambda.gt11 libraries, using labeled polypeptide or a protein fusion protein, e.g., a protein polypeptide or domain fused to a marker such as an enzyme, fluorescent dye, a
10 luminescent protein, or to an IgFc domain.

- Another embodiment of the invention relates to compositions and methods using the protein of the invention or part thereof to modulate actin organization and related cytoskeletal protein organization in cells and in particular, cells of the CNS. Compositions containing the protein of the invention and fragments thereof may be therapeutic and used to treat a variety of neuronal
15 disorders. An additional embodiment of the invention relates to the administration of a pharmaceutical or sterile composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed. The composition may be delivered via a variety of different routes.

- In yet another embodiment, an antagonist of the protein of the invention may be
20 administered to a subject to treat or prevent a neuronal disorder. Such a disorder may include, but is not limited to, akathisia, Alzheimer's disease, amnesia, amyotrophic lateral sclerosis, bipolar disorder, catatonia, cerebral neoplasms, dementia, depression, diabetic neuropathy, Down's syndrome, tardive dyskinesia, dystonias, epilepsy, Huntington's disease, peripheral neuropathy, multiple sclerosis, neurofibromatosis, Parkinson's disease, paranoid psychoses, postherpetic
25 neuralgia, schizophrenia, and Tourette's disorder. In one aspect, an antibody which specifically binds the protein of the invention may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express the protein of the invention.

- Another embodiment of the invention encompasses DNA sequences which encode the
30 proteins of the invention that may be derived through biological or synthetic chemistry processes. Polynucleotides sequences capable of hybridizing to the cDNA sequences of SEQ ID NOs: 144 and 175 are also included in the scope of the invention.

- Further included in the invention are the polypeptides encoded by the human cDNA of clones 105-021-3-0-C3-CS and 188-31-1-0-E6-CS. The polypeptides of SEQ ID NOs: 385 and 416
35 may be interchanged with the corresponding polypeptides encoded by the human cDNA of clones 105-021-3-0-C3-CS and 188-31-1-0-E6-CS. Further included in the invention are polynucleotides

encoding said polypeptides. Preferred polynucleotides are those of SEQ ID NOs: 144 and 175 and of the human cDNA of clones 105-021-3-0-C3-CS and 188-31-1-0-E6-CS.

Another embodiment of the invention to methods of using the nucleotidic sequence or part thereof of invention to search homologs to the protein of invention. The sequence can be used as
5 template of PCR reactions, allowing the detection/quantification of the protein of invention or part of thereof or the homologs. The complementary sequence or part of thereof may be used as hybridization probes to detect/quantify the transcription level, as well in *in vitro* level as the cellular level. In particular, such probes may be used in a diagnostic context, for example in the cellular or tissue *in situ* hybridization.

10 Another embodiment of the invention to methods of using the nucleotidic sequence or part thereof of invention to design antisense oligonucleotides to modulate the *in vitro* or *in vivo* expression of the protein or the part or thereof gene expression. This may be useful in therapeutic area of diseases listed above, particularly in the context where the protein of invention is expressed in abnormally high level.

15 In a further embodiment of the invention, the protein of invention or portions thereof are used to produce specific antibodies. These antibodies may have applications in the diagnostics, the purification of the protein of invention or part of thereof or a homolog. They may also help to visualize the locations of the proteins associated to the protein of invention in the cell, in particular for highlighting structure-related proteins. The methods described herein in which protein
20 antibodies are employed may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one specific cDNA of SEQ ID NO. or antibody reagent described herein, which may be conveniently used, for example, in clinical settings, to diagnose patients exhibiting symptoms of various CNS disorders.

In still another preferred embodiment, the present inventions relates to methods of using the
25 protein of the invention or part thereof in gene therapy, particularly in the diseases involving the CNS, particularly in the context where the protein of the invention is expressed in abnormally low level. Gene therapy is a potential therapeutic approach in which normal copies of the cDNAs of SEQ ID NOs: 144 and 175 may be introduced into subjects to successfully code for normal protein in several different affected cell types.

30 Another aspect of the present invention includes a formulation comprising the protein of the invention or part thereof and a pharmaceutically or physiologically acceptable carrier. A formulation of the present invention comprises a combination of one or more peptides as described herein, or mimetopes thereof; a combination of antibodies as described herein, or mimetopes thereof; or a combination of antibodies and peptides as described herein, or mimetopes thereof.
35 Such a formulation may be administered to a subject in need thereof to treat or prevent a disorder associated with decreased expression or activity of the protein. Examples of such disorders include

but are not limited to those of the CNS and other tissues where the association of actin appears to be abnormal.

Proteins of SEQ ID NO: 391, 393, 405 and 407 (internal designations 145-52-2-0-D12-CS, 145-7-3-0-D3-CS, 174-17-1-0-D6-CS and 174-38-4-0-D11-CS respectively)

5 The cluster of four proteins (SEQ ID NOs: 391, 393, 405 and 407) encoded by the cDNAs of SEQ ID NOs: 150, 152, 164 and 166 respectively exhibit very strong homology to claudin-8, a member of PMP22-Claudin family (PF00822). SEQ ID NO: 405 (174-17-1-0-D6-CS) shares a high degree of identity with human claudin-8. SEQ ID NO: 393 contains two amino acid substitutions as compared to human claudin-8 (T129A and S151P); thus, it appears to be a polymorphic variant of claudin-8. SEQ
10 ID Nos: 393 and 405 contain four membrane spanning segments.

 SEQ ID NOs: 391 and 407 are polymorphic forms of claudin-8. The protein of SEQ ID NO: 391 (145-52-2-0-D12-CS) is 162 amino-acids long, contains three theoretical membrane spanning segments, and shows three amino acid substitutions as compared to the previously identified claudin-8 protein (R31I, S151P and E162). The protein of SEQ ID NO: 407 (174-38-4-0-D11-CS) is 43 amino-
15 acids long. SEQ ID NO: 407 contains a stop codon at position 44 and contains no apparent membrane spanning segments.

 The Claudin family of proteins comprises more than twenty (20) small glycoproteins with four predicted transmembrane domains. The tissue distribution pattern of claudins varies significantly, depending on claudin species. Many have been identified as components of tight junction (TJ) strands
20 which contribute in regulation of cell polarity and permeability. Polarized epithelial and endothelial cells form barriers that separate biological compartments and regulate homeostasis. The tight junction (TJ) is a specialized membrane domain at the most apical region of polarized epithelial and endothelial cells that not only creates a primary barrier to prevent paracellular transport of solutes (barrier function) but also restricts the lateral diffusion of membrane lipids and proteins to maintain the cellular polarity
25 (fence function). Tight junctions appear to represent a continuous network of interconnected rows of intramembranous particles that appear as strands with complementary grooves. The TJ-specific integral membrane proteins, i.e. the components of TJ strands, occludins and claudins, were only recently identified.

 Claudin-1 and -2 have the ability to induce the formation of networks of strands/grooves at
30 cell-cell contact sites when introduced into fibroblasts lacking TJs. Occludin induces only a small number of short strands at cell-cell contact sites in fibroblasts, thus, it is an accessory protein in terms of TJ strand formation. Claudin transfection experiments in fibroblasts revealed the TJ strand itself can be formed without occludin (Saitou M et al. J. Cell Biol. 141: 397-408 (1998), Furuse M et al. J. Cell Biol. 143: 391-401 (1998)).

35 Initially several members of the claudin family were reported (RVPI, Clostridium perfringens enterotoxin receptor (CPE-R), and TMVCF (transmembrane protein deleted in Velo-cardio-facial

syndrome)), but their physiological functions were not determined. After the identification of claudin-1 and -2 as novel components of TJ strands (Furuse, M. et al. *J. Cell Biol.* 141, 1539-1550 (1998)), CPE-R was shown to remove specific claudins from TJs. In its presence, TJ strands in C3L cells gradually disintegrate and the number of TJ strands and the complexity of their network decreases markedly
5 (Sonoda N et al. *J Cell Biol* 147(1):195-204 (1999)). In distal tubules of the kidney, claudin-4 (CPE-R) and claudin-8 were co-localized with occludin at their junctional complex region. In liver, claudin-3 and occludin were co-localized along bile canaliculi and TJ strands were labeled heavily and specifically with anti-claudin-3 Ab (Morita et al. *PNAS* 96 (2): 511-516 (1999)). The claudins have been shown to create the paracellular diffusion barrier and, surprisingly, they may also confer channel-
10 like selectivity for passage of solutes through the tissue barrier (Anderson JM and Christina M. Van Itallie CM, *Current Biology* 9:R922-R924 (1999)).

The existence of the claudin multigene family as well as the tissue distribution pattern of each claudin species suggests that similar complexity can be expected in TJs and contributes to the generation of functional diversity of TJs in vivo. More than two distinct claudins are co-expressed in
15 single epithelial cell. Claudins interact between each of the paired strands in a heterophilic manner and distinct claudins are (except in some combinations) co-incorporated into individual TJ strands (Furuse M et al. *J Cell Biol* 147(4):891-903 (1999)).

Several claudins have been shown to be expression markers of malignant cells. For example, SEMP1 (senescence-associated epithelial membrane protein) is expressed in normal human tissues,
20 including adult and fetal liver, pancreas, placenta, adrenals, prostate and ovary; however, SEMP1 is expressed at low or undetectable levels in a number of human breast cancer cell lines (Swisshelm K et al. *Gene* 226:285-295 (1999)). Another member of the claudin family was found to be exclusively expressed in MCF-7ADR mammary carcinoma cells. MCF-7ADR carcinoma cells are estradiol-independent for growth, estrogen-receptor negative, tamoxifen resistant, vimentin positive and invasive
25 in vitro and in vivo (Schiemann S et al., *Anticancer Res* 17(1A):13-20 (1997)). Further, down regulation of the expression of claudin-1 has been associated with oncogenesis in rat salivary gland epithelium cells (Li D and Mrsny RJ *J Cell Biol* 148(4):791-800 (2000)).

The increase in microvascular permeability in human gliomas, contributing to clinically severe symptoms of brain edema, appears to be the result of a dysregulation of junctional proteins. Increased
30 TJ permeability of the colon epithelium, and consequently a decrease in epithelial barrier function, precedes the development of colon tumors, including carcinomas and adenomatous polyps (Soler AP et al. *Carcinogenesis* 20(8):1425-1431 (1999)). Studies of the interendothelial junctions in tumor microvessels of human glioblastoma multiforme show that the expression of claudin-1 is lost in the majority of tumor microvessels, whereas claudin-5 is significantly down-regulated only in hyperplastic
35 vessels. A relationship between claudin-1 suppression and the alteration of tight junction morphology is likely to correlate with the increase of endothelial permeability (Liebner S et al. *Acta Neuropathol* (Berl) 100(3):323-331 (2000)).

The human phenotype of mutations in claudin-16 suggests that it creates a channel allowing magnesium to diffuse through renal tight junctions. Similarly, a mouse knockout of claudin-11 reveals its role in formation of tight junctions in myelin and between Sertoli cells in testis (Mitic LL et al. Am J Physiol Gastrointest Liver Physiol 279(2):G250-254 (2000)).

- 5 Opening of TJs by environmental proteinases may be the initial step in the development of asthma to a variety of allergens. The lung epithelium forms a barrier that allergens must cross before they can cause sensitization. The cysteine proteinase allergen Der p 1 from fecal pellets of *Dermatophagoides pteronyssinus* (the house dust mite (HDM)) causes disruption of intercellular tight junctions (TJs), which are the principal components of the epithelial paracellular permeability barrier.
- 10 TJ breakdown nonspecifically increases epithelial permeability, allowing Der p 1 to cross the epithelial barrier. Putative Der p 1 cleavage sites were found in peptides from an extracellular domain of claudin-1. House dust mite (HDM) allergens are important factors in the increasing prevalence of asthma (Wan H et al. J Clin Invest 104(1):123-33 (1999)).

- In many intestinal and systemic diseases, intestinal barrier damage is marked by changes in
- 15 intestinal permeability which are, in turn, related to alteration in tight junction function (Gasbarrini G, Montalto Mital J Gastroenterol Hepatol 31(6):481-488 (1999)). Permeability of the tight junctions can be modified by bacterial toxins, cytokines, hormones and drugs. Oligodendrocyte-specific protein (OSP/claudin-11), found in CNS myelin, appears to be a promising candidate for auto-antigenic involvement in autoimmune demyelinating disease. The presence of anti-OSP Abs in the cerebrospinal
- 20 fluid was reported for relapsing-remitting multiple sclerosis (MS). Murine OSP peptides elicit clinical experimental autoimmune encephalomyelitis in animal models for MS and induces mononuclear cell infiltrates and focal demyelination. Also OSP peptides elicit robust proliferative responses in T cells (Stevens DB et al. J Immunol 162:7501-7509 (1999)). OSP/claudin-11 appears to modulate proliferation and migration of oligodendrocytes, presumably through the membrane interactions at tight
- 25 junctions and with the extracellular matrix (Bronstein JM et al. J Neurosci Res 59(6):706-711 (2000)). Recently claudin-11 has been shown to play a key role in the formation of hematotesticular barrier; it is regulated by FS hormone and by cytokines in early fetal and postnatal development in Sertoli cells (Hellani A. et al. Endocrinology 141: 3012-3019 (2000)).

- SEQ ID NOs: 391, 393, 405 and 407 are new human proteins having biological activities
- 30 described for claudins. Nucleic acids encoding the proteins of interest are over represented in fetal kidney and in salivary gland. The subject invention provides polynucleotides encoding the proteins of SEQ ID Nos: 391, 393, 405 and 407. In one embodiment, the polypeptides of SEQ ID NOs: 391, 393, 405 and 407 are interchanged by the polypeptides encoded by clones 145-52-2-0-D12-CS, 145-7-3-0-D3-CS, 174-17-1-0-D6-CS and 174-38-4-0-D11-CS. Also provided are use of these proteins,
- 35 fragments, derivatives thereof (and related polynucleotides) for the diagnosis, treatment, or prevention of tumors and another diseases, including disorders associated with altered epithelial function. The invention also encompasses possible variants of the proteins of interest which have at least about 80%,

more preferably at least about 90%, and most preferably at least about 95% amino acid sequence identity to the amino acid sequence, provide the variants have at least one of the functional or structural characteristics of the identified claudin-like proteins.

In one embodiment of the subject invention, the proteins of interest, or biologically active fragments or variants thereof, may be administered to a subject to treat or prevent disorders of salivary gland, kidney and prostate. The subject invention also provides therapeutic regimens for the treatment of epithelial dysfunction and cancer.

The disorders which may be treated in accordance with the subject invention include, but are not limited to, asthma, eczema, atopic dermatitis, contact dermatitis, stasis dermatitis, seborrheic dermatitis, psoriasis, lichen planus, pityriasis rosea, acne vulgaris, acne rosacea, pemphigus vulgaris, pemphigus foliaceus, paraneoplastic pemphigus, bullous pemphigoid, herpes gestationis, dermatitis herpetiformis, linear IgA disease, epidermolysis bullosa acquisita, dermatomyositis, lupus erythematosus, scleroderma, and morphea; gastritis, peptic ulcers, cholelithiasis, cholecystitis, pancreatitis, cirrhosis, ulcerative colitis, Crohn's disease, and irritable bowel syndrome; Addison's disease, Lowe's syndrome, glomerulonephritis, chronic glomerulonephritis, tubulointerstitial nephritis, inherited X-linked nephrogenic diabetes insipidus, autosomal dominant polycystic kidney disease, autoimmune demyelinating disease, multiple sclerosis, glioma, and other tumors.

A further aspect of the invention provides a method for treating these and/or other pathological states by administering, to a patient, a therapeutically effective amount of one or more of the proteins of interest. The proteins of interest may, optionally, be simultaneously or sequentially administered in conjunction with cytokines and/or interleukins which have been shown to improve claudin expression.

In another embodiment, a vector capable of driving expression of one or more of the proteins of interest, or a biologically active fragment or variant thereof, may be administered to a subject to treat or prevent an epithelial permeability disorder including, but not limited to aforementioned disorders.

Another embodiment of the subject invention provides compositions and methods of treating, or reducing the incidence of, asthma comprising the administration of therapeutically effective amounts of the proteins of the subject invention. In one embodiment, purified fragments, or synthetically modified peptides, derived from the extracellular domains of the proteins of interest may be administered in the therapeutic regimen. The peptides, containing the putative cleavage sites for environmental allergen proteinases, may be administered in amounts which competitively inhibit the proteinase activity of the allergen. The peptides may be designed to bind allergen, optionally in an irreversible manner, and inhibit proteinase activity.

The negative effects of the usual preservation solutions on epithelial and endothelial permeability in organs to be transplanted are generally known (Trocha S.D. et al. Ann.Surg. 230: 105-113 (1999)). Increases in permeability leads to tissue injury and edema. Disorganization of tight junctional proteins appears to be responsible for the observed tissue injury and edema. Thus, in another embodiment, purified proteins of interest, or variants and/or biologically active fragments thereof, may

be added in organ preservation solutions to maintain the content and integrity of tight junctions in organs.

In another embodiment, the subject invention provides methods of producing "bioartificial" epithelia from non-epithelial cells. The "bioartificial" epithelia produced according to the invention may be used for reconstructive surgical procedures, for treating of disorders related to epithelial loss (for hereditary, traumatic or oncological reasons) or for another therapeutic purposes (e.g., burn treatments). "Bioartificial" epithelial cells can be obtained by transfection and remodeling of the autologous patient cells not affected by any of the aforementioned disorders. The use of autologous cells in the preparation of the "bioartificial" epithelial cells of the invention in methods of treating disorders, conditions, or diseases associated with the loss of epithelial cells reduces or eliminates the risk of tissue rejection typically observed in transplantation methodologies. Methods of bioartificial tissue engineering are generally known to those skilled in the art (for a review, see Machens H.G. et al. Cells Tissues Organs 167: 88-94 (2000)).

In another embodiment of the subject invention provides antibodies which specifically bind to the proteins of SEQ ID Nos: 391, 393, 405 and 407. The antibodies may also specifically bind to fragments or variants of the proteins described in SEQ ID Nos: 391, 393, 405 and 407. The antibodies of the invention may be used to detect the protein of interest in human body fluids, extracts of cells or tissue extracts. The detection assays may be used for epithelial cancer prognosis and for the diagnosis of disorders. The assays may also be used to monitor patients being treated with the proteins of interest.

In another embodiment, the polynucleotide sequences, or fragments of said polynucleotide sequences, encoding the proteins of interest may be used for the identification or diagnosis of a disorder associated with expression of the proteins of SEQ ID Nos: 391, 393, 405 and 407. Hybridization assays which allow for the detection of polynucleotide sequences of the invention are well known to the skilled artisan. These assays include, and are not limited to, Northern blots, Southern blots, and PCR methodologies.

Another embodiment of the invention provides the proteins of SEQ ID Nos: 391, 393, 405 and 407, variants, immunogenic fragments, or biologically active fragments of said proteins for screening libraries of compounds in any of a variety of drug screening techniques. The proteins of SEQ ID Nos: 391, 393, 405 and 407, variants, immunogenic fragments, or biologically active fragments of said proteins employed in such screening may be free in solution, affixed to a solid support, recombinantly expressed on, or chemically attached to, a cell surface, or located intracellularly. The formation of binding complexes between the protein of interest and the agent being tested may be measured by methods well known to those skilled in the art.

Yet another embodiment of the invention provides methods of screening compounds which modulate epithelial permeability Polynucleotides encoding the proteins of SEQ ID Nos: 391, 393, 405 and 407, variants, immunogenic fragments, or biologically active fragments of said proteins, may be

recombinantly expressed in cells typically lacking TJs according to methods discussed supra. These cells may then be used to assess therapeutic modulators (based, for example, on CPE –like compounds) for the ability to increase or decrease epithelial cell permeability. Compounds identified in these modulator screen assays may then be used in therapeutic protocols to adjust epithelial cell permeability as desired by the practitioner.

The intestinal epithelium is a major barrier to the absorption of hydrophilic drugs. The presence of intercellular junctional complexes, particularly the tight junctions, renders the epithelium impervious to hydrophilic drugs, which cannot diffuse across the cells through the lipid bilayer of the cell membranes (Ward PD et al. *Pharmaceutical Science and Technology Today* 3:10:346-358 (2000)).

Therefore, in another embodiment of the subject invention the proteins of SEQ ID Nos: A, B, C, or D, variants, or biologically active fragments of said proteins and their molecular partners can be used for the rational design of compounds that can effectively and safely increase paracellular permeability for selected drugs. For example, polynucleotides encoding the proteins of interest or any fragment or derivatives thereof, may be used for these purposes. In one aspect, the complement of the polynucleotide encoding the protein of interest may be used in situations in which it would be desirable to block the transcription of the mRNA encoding the proteins of interest, especially for temporally increasing epithelial permeability (useful for drug delivery). Alternatively, sense or antisense oligonucleotides may be designed from various locations along the coding or control regions of polynucleotide sequences encoding the proteins of SEQ ID Nos: 391, 393, 405 and 407, as well as variants, or biologically active fragments of said proteins to control expression of the proteins. Methods of producing and using sense and antisense oligonucleotides are well known to those skilled in the art.

Claudins are unique proteins with specific protein-binding properties. Therefore, in another preferred embodiment, the proteins of SEQ ID Nos: 391, 393, 405 and 407, variants, or biologically active fragments of said proteins may be used as a component of drug delivery vehicles such as colloids or liposomes. The proteins of the proteins of SEQ ID Nos: 391, 393, 405 and 407, variants, or biologically active fragments of said proteins may be incorporated into the lipid membranes of liposomes and can serve as specific targeting agents which bind the specific epithelial targets and facilitate targeted epithelium drug delivery. The methods of design of such type of drug delivery systems is known by those skilled in the art (Smith H.J. *Introduction to the principles of drug design and action*, 3rd ed. (1998)). Alternatively, active agents, such as chemotherapeutic agents, radioisotopes, prodrugs, may be directly attached, recombinantly or chemically, to the proteins of SEQ ID NOs: 391, 393, 405 and 407, variants, or biologically active fragments of said proteins and used in therapeutic regimens.

Proteins of SEQ ID Nos: 278, 282 and 300 (internal designations 160-37-2-0-H7-CS, 174-33-3-0-F6-CS, 184-4-1-0-A11-CS respectively)

The protein of SEQ ID No: 278 (and the corresponding allelic variants 282 and 300) encoded by the cDNA SEQ ID No: 37 (41, 59 respectively) shows homology to a human transmembrane protein (HTMN-23, Genseq accession number Y57899). The protein of SEQ ID No: 278 (and the corresponding polymorphic variants 282 and 300) overexpressed in salivary gland contains 9 potential transmembrane segments from positions 85 to 105, 116 to 136, 164 to 184, 187 to 207, 332 to 352, 376 to 396, 404 to 424, 465 to 485 and 499 to 519, thus displaying characteristic features of type III transmembrane proteins (Singer, Annu. Rev. Cell Biol., 6:247-296, 1990). Furthermore, a predicted localisation in the endoplasmic reticulum (ER) is found for the protein of the invention with the software psort.

The normal functioning of the eukaryotic cell requires that all the newly synthesized proteins be correctly folded, modified, and delivered to specific intra- and extracellular sites. Newly synthesized membrane and secretory proteins enter a cellular sorting and distribution network during or immediately after synthesis and are routed to specific locations inside and outside the cell. The initial compartment in this process is the endoplasmic reticulum (ER) where proteins undergo modifications such as glycosylation, disulfide bond formation, and assembly into oligomers. The modified proteins are then transported through a series of membrane-bound compartments which include the various cisternae of Golgi complex where further carbohydrate modifications occur. Transport between compartments occurs by means of vesicles that bud and fuse in a manner specific to the type of protein being transported. Once within the secretory pathway, proteins do not have to cross a membrane to reach the cell surface. Disruptions in the cellular secretory pathway have been implicated in several human diseases. In familial hypercholesterolemia the low density lipoprotein receptors remain in the ER rather than moving to the cell surface (Pathak, et al., J. Cell Biol., 106:1831-1841, 1988). Altered transport and processing of the beta-amyloid precursor protein (betaAPP) involves the putative vesicle transport protein prenesilin, and may play a role in early-onset Alzheimer disease (Levy-Lahad et al., Science, 269:973-977, 1995). Changes in the ER-derived calcium homeostasis have been associated with diseases such as cardiomyopathy, cardiac hypertrophy, myotonic dystrophy, Brody disease, Smith-McCort dysplasia and diabetes mellitus.

It is believed that the protein of the invention represents a new ER integral transmembrane protein. This protein plays probably a role in post-translational modifications of secreted and membrane proteins. Its dysregulated expression may be linked to disorders such as the above referred diseases. Preferred polypeptides of the invention are polypeptides comprising the amino acids of SEQ ID Nos: 278, 282 and 300 from positions 85 to 105, 116 to 136, 164 to 184, 187 to 207, 332 to 352, 376 to 396, 404 to 424, 465 to 485 and 499 to 519. Other preferred polypeptides of the invention are fragments of SEQ ID Nos: 278, 282 and 300 having any of the biological activity described herein.

One object of the present invention are compositions and methods of targeting heterologous polypeptides to the endoplasmic reticulum by recombinantly or chemically fusing a fragment of the proteins of the invention to an heterologous polypeptide. Preferred fragments are any fragments of the proteins of the invention, or part thereof, that may contain targeting signals for the endoplasmic
5 reticulum such as those described in Pidoux AL, Armstrong EMBO J 1992 Apr;11(4):1583-91; Munro S, Pelham HR Cell 1987 Mar 13;48(5):899-907; Pelham HR Trends Biochem Sci 1990 Dec;15(12):483-6.

In another embodiment, the invention relates to methods and compositions using the protein of the invention or part thereof as marker proteins to selectively identify tissues, preferably salivary
10 glands. For example, the proteins of the invention or part thereof may be used to synthesize specific antibodies using any techniques known to those skilled in the art including those described therein. Such tissue-specific antibodies may then be used to identify tissues of unknown origin, for example, forensic samples, differentiated tumor tissue that has metastasized to foreign bodily sites, or to differentiate different tissue types in a tissue cross-section using immunochemistry.

15 Moreover, antibodies to the proteins of the invention or parts thereof may be used for the detection of endoplasmic reticulum in immunochelistry for example using any techniques known to those skilled in the art including those described herein.

Protein of Seq Id No: 281 (174-10-2-F8-CS)

The protein of SEQ ID No: 281 is homologous to PET117 (SwissProt ID: Q02771). MTC
20 is overexpressed in the brain, dystrophic muscle, fetal liver, placenta and salivary glands.

The protein of the invention, herein named MTC, presents a certain homology with the yeast PET117 protein precursor (22 % identical amino acids, 39% positive amino acids when aligned by BLASTP 2.0.9). MTC appears to be a novel member of the PET family.

Cytochrome *c* oxidase (complex IV), an enzyme complex located in the mitochondrial inner
25 membrane, is the terminal member of the mitochondrial electron transport chain. The oxidation reaction catalyzed by cytochrome *c* oxidase is exergonic and is coupled to the translocation of protons across the membrane. This reaction provides the energy needed to drive the synthesis of ATP by the mitochondrial oxidative phosphorylation system and is essential for respiratory metabolism in aerobic eukaryotes. Cytochrome *c* oxidase is made up of as many as 13 non-
30 identical protein subunits, of which 3 are encoded by the mitochondrial genome, and contains several prosthetic groups (including heme groups *a* and *a₃*).

The composition of cytochrome *c* oxidase requires that synthesis and assembly of a functional enzyme complex occur in several distinct steps, including: 1) synthesis of the protein subunits, 2) transport of the subunits from their site of synthesis to their site of function in the
35 mitochondrial inner membrane, 3) synthesis of hemes *a* and *a₃* and, 4) assembly of the subunits with each other and with the prosthetic group. A number of "accessory" genes (e.g., not encoding

protein subunits of the final assembled cytochrome *c* oxidase complex) are required for the production of functional cytochrome *c* oxidase (McEwen JE et al. – J Biol Chem – 1986, 261(25):11872-9). Some of these are required for the expression of mitochondrial-encoded cytochrome *c* oxidase subunits, while others are needed for the proper assembly of active

5 cytochrome *c* oxidase.

The nuclear genes PET117 and PET 191 belong to this class of “accessory genes” required for the assembly of active mitochondrial cytochrome *c* oxidase (McEwen JE et al. – Curr Genet. – 1993, 23(1):9-14). The role of PET genes and the proteins that they encode remains obscure, although mutation experiments in *S cerevisiae* have clearly shown that they are essential for the

10 production of active cytochrome *c* oxidase (McEwen JE et al. – J Biol Chem – 1986, 261(25):11872-9) (McEwen JE et al. – Curr Genet. – 1993, 23(1):9-14).

One aspect of the subject invention provides to compositions and methods of using the nucleotide sequence of SEQ ID No: 40, or its complement, in molecular biology techniques. In one embodiment, the MTC2 sequence is encoded by clone 174-10-2-0-F8-CS. References to a

15 polynucleotide of SEQ ID NO: 40 and polypeptide of SEQ ID NO: 281 are interchangeable with the corresponding polynucleotides of the human cDNA of clone 174-10-2-0-F8-CS and polypeptides encoded thereby. These techniques include, but are not limited to: PCR; production of recombinant MTC, or biologically active fragments thereof, generating antisense RNA and DNA, their chemical analogs and the like; hybridization probes; and chromosome gene mapping.

20 As is apparent to one skilled in the art, all of the non-limiting techniques listed above can be practiced with fragments of the *mtc2* gene. Given the well-known nature of these techniques, the skilled artisan will be able to select an appropriate length of the *mtc2* polynucleotide for use in the techniques. For recombinant expression of protein, a preferred embodiment provides the full length MTC2 gene in an expression vector.

25 For example, nucleotide sequence of SEQ ID No: 40 or its complement can be used to generate hybridization probes for mapping the naturally occurring genomic sequence. The sequence can be mapped to a particular chromosome or to a specific region of the chromosome using well-known techniques. These include in situ hybridization to chromosomal spreads, flow-sorted chromosomal preparations, or artificial chromosome constructions such as yeast artificial

30 chromosomes, bacterial artificial chromosomes, bacterial P1 constructions or single chromosome cDNA libraries as reviewed in Price (Price CM – Blood Rev. – 1993, 7(2):127-34) and Trask B (Trask BJ – Trends Genet. – 1991, 7(5):149-54).

In situ hybridization of chromosomal preparations and physical mapping techniques such as linkage analysis using established chromosomal markers are invaluable in extending genetic maps

35 that provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once a disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, any sequences mapping to that area can represent

associated or regulatory genes for further investigation. The nucleotide sequence of the present invention can also be used to detect differences in the chromosomal location due to translocation, inversion, etc. among normal, carrier or affected individuals.

The subject invention also provides methods of using MTC polypeptides and
5 polynucleotides encoding said polypeptides in preventing or reducing the incidence of apoptosis in cells. Dysfunctions in the mitochondrial electron transport chain result in cellular apoptosis or necrosis. In one embodiment, MTC is added to an *in vitro* culture of mammalian cells in an amount effective to reduce apoptosis. In another embodiment, cells are transfected with vectors comprising MTC polynucleotides which cause the expression of MTC peptides. MTC used in these
10 embodiments can, optionally, contain mitochondrial targeting sequences. In another embodiment, MTC or MTC2 are encoded by clone 174-10-2-0-F8-CS.

In another embodiment, MTC polypeptides and polynucleotides encoding said polypeptides can be used in the diagnosis, treatment and/or prophylaxis of disorders associated with apoptosis or impairment of the mitochondrial respiratory electron transport chain. Polynucleotides can also be
15 used in antisense protocols for certain disorders to impair the function of the mitochondrial electron transport chain. These disorders include, but are not limited to, immune deficiency syndromes (including AIDS); type I diabetes; pathogenic infections; cardiovascular and neurological injury; alopecia; aging; degenerative diseases such as Alzheimer's Disease, Parkinson's Disease, Huntington's disease; dystonia; Leber's hereditary optic neuropathy; schizophrenia; neonatal hepatic
20 failure and ketoacidotic coma necrosis; and myodegenerative disorders such as "mitochondrial encephalopathy, lactic acidosis, and stroke" (MELAS), "myoclonic epilepsy ragged red fiber syndrome" (MERRF); mitochondriocytopathies, Leigh syndrome, fatal infantile cardioencephalomyopathy, ataxia; encephalopathies, aging, neurodegenerative diseases, myopathies, and cancers. As would be apparent to the routineer, these methods can be practiced
25 with full length MTC polypeptides and polynucleotides encoding said polypeptides as well as biologically active fragments of the same which retain biological activity.

For diagnostic purposes, the expression of the protein of the invention could be investigated using any of the Northern blotting, RT-PCR or immunoblotting methods well known to those skilled in the art. For prophylaxis and/or treatment purposes SEQ ID No: 40, its complement, or
30 fragments of either, can be used to enhance electron transport and increase energy delivery using any of the gene therapy methods known to those skilled in the art. Likewise, SEQ ID NO: MTC2, its complement, and fragments of either can be used to impair electron transport and decrease energy delivery using any of the antisense methodologies known to those skilled in the art.

Protein of SEQ ID NO:392 (145-7-2-0-G5-CS)

The protein of SEQ ID No:392, encoded by the cDNA of SEQ ID No:151, is homologous to Unc-18 proteins, also known as the STXBP or Sec-1 family. The protein of the invention is strongly expressed in the fetal kidney.

5 Amino acids 89 to 107 of the protein of the invention present the EMotif signature for proteins of the Sec-1 family (BlocksPlus PF00995). Furthermore, BLAST analysis (BLASTP version 2.0.9) of the amino acid sequence of the invention reveals that it is homologous to a number of proteins belonging to the Unc-18/Sec-1 family. Preferred polypeptides of the invention are those that comprise amino acids 94, 95, and/or 100, which are conserved in more than 80% of the Sec-1
10 family members; and/or amino acids 43, 89, and/or 97, which are conserved in more than 60% of Sec-1 family members. Other preferred polypeptides of the invention are any fragment of SEQ ID NO:392 having any of the biological activities described herein.

The normal function and organization of eukaryotic cells is dependent on transport of various vesicles that selectively shuttle membrane and cargo between distinct compartments of the
15 secretory and endocytotic pathways. A number of key proteins involved in membrane targeting and exocytosis have been identified, and a fundamental set of interactions has been defined and placed into a model called the SNARE (Soluble N-ethylmaleimide-sensitive Attachment protein REceptor) hypothesis (Rothman J – Nature – 1994, 372: p55-63). According to the SNARE hypothesis, vesicles dock to a target membrane through the interaction of complementary sets of vesicular (v-
20 SNARE) and target (t-SNARE) membrane proteins. Our understanding of vesicle trafficking has, to a large extent, been facilitated by characterization of synaptic vesicles in neurons. In synaptic vesicle exocytosis, the vesicular protein synaptobrevin (also called Vesicle-Associated Membrane Protein; VAMP) is the v-SNARE, and the plasma membrane-associated protein SNAP-25 (Synaptosomal-Associated Protein of 25 kDa) and syntaxin 1 function as t-SNARE. Formation of
25 the SNARE complex (or core complex) is followed by recruitment of the cytosolic proteins alpha, beta and gamma SNAP (Soluble N-ethylmaleimide-sensitive Attachment Protein) and NSF (N-ethylmaleimide-Sensitive Factor), which are required for membrane fusion. Proteins from two gene families have been identified as key regulators of SNARE complex assembly. These include members of the small GTP-binding family (e.g. Rabs) and the Sec-1 family. The Sec-1 gene is one
30 of ten genes identified as essential for the final stages of protein secretion in yeast (*S. cerevisiae*). Sec-1 homologues have been identified in the nervous system of *C. elegans* (Unc-18), *D. melanogaster* (Rop) and mammals. In mammals, the protein has been termed Mammalian homologue of the Unc-18 gene (Munc-18), rbSec1 (Rat Brain Sec1) or n-Sec1 (neural-specific Sec1).

35 Sec-1-related proteins are involved in the processes of vesicle targeting, docking and/or fusion. Sec-1-related proteins interact directly with the t-SNARE syntaxin, and Munc-18 has been found to interact with syntaxin isoforms 1a, 2 and 3. However, Munc-18 has not been found to be

part of the 20S SNARE/SNAP/NSF protein complex. In vitro, the binding of Munc-18 to syntaxin inhibits the interaction of syntaxin with VAMP and SNAP-25 as well as SNAP-23 (a homologue of SNAP-25) and thereby negatively regulates the formation of the synaptic SNARE fusion complex. In agreement with a negative regulatory role of Sec-1/Munc-18 proteins in neurotransmitter release

5 are results showing that microinjections of Sec-1 into the presynaptic terminal of the giant squid synapse inhibits evoked transmitter release (Dresbach T. et al. – J Neurosci. – 1998, 18: p2923-2932). Furthermore, overexpression of Rop, Unc-18, Sec-1 and Munc-18 all result in phenotypes associated with a complete block in neurotransmitter release and/or secretion (Hosono R. et al. – J Neurochem – 1992; 58: p1517-1525; Harrison S. et al. – Neuron – 1994; 13: p555-566; Novick P.

10 et al. – Cell – 1981; 25: p461-469; verhage M. et al. – Science – 2000; 287: p864-869). Point mutation experiments involving the Rop gene suggest that Rop is a rate-limiting regulator of exocytosis that performs both stimulatory and inhibitory functions in neurotransmission (Wu M. et al. – EMBO J – 1998; 17: p127-139). The reduction in neurotransmitter release seen after both overexpression of Munc-18 and mutations in Munc-18 homologues indicates that Sec-1 proteins not

15 only sequester syntaxins from other proteins but also assist the syntaxins in adopting a functional conformation or facilitate interactions between syntaxins and other proteins by a chaperone-like action. The necessity of Sec1-related proteins is believed to result, in part, from their direct and high affinity interaction with members of the t-SNARE family of syntaxin proteins and from the control by this complex of a v- and t-SNARE protein interaction required for vesicle fusion.

20 The SNARE mechanism of exocytosis appears to be conserved both evolutionarily (most of the components have homologues in species from yeast to mammals) and functionally (each of the principal components are members of multigene families). This latter point is supported by work showing that components of this pathway are found in different cell types (neurons, neutrophils and pancreatic beta-cells) (brumell J. et al. – J Immunol – 1995; 155: p5750-5759; Zhang W et al. –

25 J Biol Chem. – 2000 Oct 6, electronic publication).

It is believed that the protein of SEQ ID NO:392 is a member of the Unc-18/Sec-1 family, and thus plays a key role in the regulation of various processes including vesicle targeting, docking and fusion.

One embodiment of the present invention relates to the use of the protein of SEQ ID

30 NO:392 or the cDNA of SEQ ID NO:151 or any part thereof to used to identify fetal kidney tissue and cells derived from this tissue, since the protein of the invention is strongly expressed in this tissue. In addition, the protein of the invention can be used to specifically label components of the secretory pathway within cells. Assays for the detection of cells expressing the protein of the invention, or part thereof, can be developed using techniques known to those skilled in the art. For

35 example, the protein of the invention, or part thereof, can be used to generate antibodies or antiserum, by techniques well known to those skilled in the art. Antibodies or antiserum can also be used for quantitative analysis or detection of the protein of the invention, by methods such as

enzyme-linked immunosorbant assays (ELISA) or by any other technique known to those skilled in the art. Another possible technique involves the use of marked syntaxins, since Sec1-related proteins are known to bind to syntaxins.

In another embodiment of the present invention, the present polynucleotides and polypeptides can be used to diagnose, treat, and/or prevent any of a large number of diseases and disorders characterized by abnormal exocytosis, such as, but not limited to: allergies including hay fever, asthma, and urticaria; neurologic disorders, a number of which result from abnormal neurotransmitter secretion (for example, depression is associated with decreased serotonin secretion); autoimmune hemolytic anemia; cancers, especially hormone-dependent cancers such as those stimulated by androgens (for example, prostate cancer) or estrogens (for example, breast cancer), leukemias or lymphomas; ulcerative colitis; type 2 diabetes, which in some cases is associated with decreased insulin secretion; proliferative granulonephritis; inflammatory bowel disease; growth failure due to decreased secretion of growth hormone; multiple sclerosis; myasthenia gravis, rheumatoid and osteoarthritis; scleroderma,; Chediak-Higashi and Sjogren's syndromes; systemic lupus erythematosus; thyroiditis; toxic shock syndrome; traumatic tissue damage; viral, bacterial, fungal and protozoal infections; and other physiologic/pathologic disorders associated with induced or otherwise abnormal vesicular trafficking.

An association between the level of expression and/or activity of the present protein with the presence or absence of any condition associated with abnormal vesicular trafficking, such as any of the above-listed disorders, can readily be assessed by detecting the level of expression or activity of the protein by, e.g., Northern blot, western blot, ELISA, or any standard in vitro or in vivo assay for protein activity, and correlating the observed level or expression or activity with the presence or absence of the disorder. For those disorders found to be positively associated with the protein of the invention, a diagnostic or screening assay can be readily developed where the detection of an elevated level of protein or protein activity is indicative of the presence of the disease, or of a propensity to develop the disease. Further, any such diseases or conditions can be treated or prevented by inhibiting the expression or activity of the protein, for example by administering to a patient suffering from the disorder any inhibitor including, but not limited to, antibodies, antisense oligonucleotides, dominant negative forms of the protein, and small molecule inhibitors of protein expression or activity. Alternatively, disorders negatively associated with the protein of the invention can be diagnosed or screened for by detecting the level of the present protein or protein activity, where a decreased level of the protein or protein activity is indicative of the presence of the disease, or of a propensity to develop the disease. Such disorders negatively associated with the protein of the invention can be treated or prevented by increasing the level of the protein or protein activity, for example by administering to a patient any of a number of agents including, but not limited to, the protein itself, a polynucleotide encoding the protein, or a heterologous compound that enhances the expression or activity of the protein.

protein of SEQ ID NO:419 (internal designation 188-9-1-0-C10-CS)

The protein of SEQ ID NO:419, highly expressed in the brain and placenta, is encoded by the cDNA of SEQ ID NO:178, is localized preferentially in the endoplasmic reticulum, and is homologous to the yeast integral membrane protein SFT2p, a member of the SNARE-related family
5 (Genbank accession number X79489). SFT2p is well conserved in *C. elegans* and in mice (accession numbers CAA93859 and AA790425 respectively), and plays an important role in the protein trafficking and fusion machinery of eukaryotic cells. The 159-amino-acid-long protein of the invention, which is similar in size and in membrane topology to the SFT2p protein, displays
10 four conserved hydrophobic stretches from positions 36 to 56, 66 to 86, 98 to 118 and 122 to 142, forming a tetra-spanning membrane protein. This topology is also found in the Got1p protein, another well-conserved SNARE related protein with similar functions to those of SFT2p protein (accession number AL010285 for *P. falciparum*, U23521 for *C. elegans*) as described in Conchon et al., EMBO J., 18(14):3934-3946 (1999).

Eukaryotic proteins are synthesized within the endoplasmic reticulum (ER), are delivered
15 from the ER to the Golgi complex for post-translational processing and sorting, and are transported from the Golgi to specific intracellular and extracellular destinations. This intracellular and extracellular movement of protein molecules is termed vesicle trafficking. Trafficking is accomplished by the packaging of protein molecules into specialized vesicles which bud from the donor organelle membrane and fuse to the target membrane (Palade, Science 189:347-358 (1975)).

20 Numerous proteins are necessary for the formation, targeting, and fusion of transport vesicles and for the proper sorting of proteins into these vesicles. The vesicle trafficking machinery includes coat proteins which promote the budding of vesicles from donor membranes, vesicle- and target-specific identifiers (v-SNAREs and t-SNAREs) which bind to each other and dock the vesicle to the target membrane (Nichols et al., Nature 387:199-202, 1997), and proteins which bind to
25 SNARE complexes and initiate fusion of the vesicle to the target membrane (SNAPs).

SFT2p is a conserved yeast protein with four transmembrane domains that is resident in punctate structures corresponding to the late Golgi compartment, and which enters presumptive retrograde intra-Golgi vesicles whose fusion depends on two t-SNARE proteins Sed5p and Sft1p (Wooding and Pelham, Mol. Biol. Cell 9:2667-2680 (1998)). Its genetic interaction with Sed5p
30 suggests that SFT2p is an additional membrane component involved in the docking or fusion process. In vivo experiments have shown that deletion of GOT1p or SFT2p alone does not affect cell growth, but repression of both of these proteins results in a significant accumulation of ER membrane, suggesting that the presence of either SFT2p or GOT1p is required for the maintenance of efficient ER-Golgi transport (Conchon et al., supra). It has also been shown that Got1p normally
35 facilitates Sed5p-dependant fusion events, while Sft2p performs a related function in the late Golgi (Conchon et al., supra).

The etiology of numerous human diseases and disorders can be attributed to defects in the trafficking of proteins to organelles or the cell surface. For example, defects in the trafficking of membrane-bound receptors and ion channels have been implicated in cystic fibrosis (cystic fibrosis transmembrane conductance regulator; CFTR), glucose-galactose malabsorption syndrome

5 (Na.sup.+ /glucose cotransporter), hypercholesterolemia (low-density lipoprotein (LDL) receptor), and forms of diabetes mellitus (insulin receptor). Abnormal hormonal secretion has been linked to disorders including diabetes insipidus (vasopressin), hyper- and hypoglycemia (insulin, glucagon), Grave's disease and goiter (thyroid hormone), and Cushing's and Addison's diseases (adrenocorticotrophic hormone; ACTH).

10 Further, cancer cells secrete excessive amounts of hormones or other biologically active peptides. Disorders related to excessive secretion of biologically active peptides by tumor cells include: fasting hypoglycemia due to increased insulin secretion from insulinoma-islet cell tumors; hypertension due to increased epinephrine and norepinephrine secreted from pheochromocytomas of the adrenal medulla and sympathetic paraganglia; and carcinoid syndrome, which includes
15 abdominal cramps, diarrhea, and valvular heart disease, caused by excessive amounts of vasoactive substances (serotonin, bradykinin, histamine, prostaglandins, and polypeptide hormones) secreted from intestinal tumors. Ectopic synthesis and secretion of biologically active peptides (peptides not expected from a tumor) includes ACTH and vasopressin in lung and pancreatic cancers; parathyroid hormone in lung and bladder cancers; calcitonin in lung and breast cancers; and thyroid-stimulating
20 hormone in medullary thyroid carcinoma.

It is believed that the protein of SEQ ID NO:419 or part thereof is an integral membrane protein of the SNARE-related family, and more presumably is the human homologue of the yeast SFT2p protein. Thus, the protein of the invention plays a role in the secretory and endocytic pathway of eukaryotic cells through fusion and transport of vesicles from the endoplasmic reticulum
25 to late Golgi cisternae. Preferred polypeptides of the invention are polypeptides comprising the amino acids of SEQ ID NO:419 of the four transmembrane domains from positions 36 to 56, 66 to 86, 98 to 118 and 122 to 142. Other preferred polypeptides of the invention are fragments of SEQ ID NO:419 having any of the biological activities described herein.

In one embodiment, the invention relates to methods and compositions using the protein of
30 the invention or part thereof as a new marker protein to selectively identify secretory and endocytic traffic, preferably in the endoplasmic reticulum and more preferably in the late Golgi cisternae. For example, the protein of the invention or part thereof may be detected using specific antibodies generated against the protein using any technique known to those skilled in the art. Such organelle-specific antibodies may then be used to identify cells with disrupted trafficking systems such as in
35 differentiated tumor cells or to differentiate specific organelle types in a cell cross-section using immunochemistry. In addition, the protein of the invention can be used to specifically identify cells of the brain and/or placenta, tissues in which the protein is overexpressed.

Another embodiment of the present invention relates to methods of targeting heterologous compounds, such as polypeptides or polynucleotides, to the endoplasmic reticulum and preferentially to late Golgi vesicles by recombinantly or chemically fusing a fragment of the protein of the invention to the heterologous polypeptide or polynucleotide. Such fusion proteins may be engineered to contain a cleavage site located between a sequence encoding the protein of the invention and the heterologous protein sequence, so that the protein of the invention may be cleaved and purified away from the heterologous moiety. Preferred fragments of the protein that can be used in such applications are the four transmembrane domains or any other fragments of the protein of the invention, or part thereof, that may contain targeting signals for ER or Golgi organelles as defined in Conchon et al., supra; Wooding and Pelham, supra. Such heterologous compounds may be targeted to the secretory pathway to modulate ER-Golgi endocytic and secretory activities. In one embodiment, the protein of the invention can be used to screen peptide libraries for inhibitors of traffic activity, as detected by the accumulation of ER membranes or Golgi vesicles as described in Conchon et al., supra.

In still another embodiment, the protein of the invention is used to diagnose, prevent and/or treat any of a number of disorders in which trafficking and/or the fusion machinery is affected, including, but not limited to, endocrine, secretory, inflammatory, and gastrointestinal disorders, such as cancer, cystic fibrosis (cystic fibrosis transmembrane conductance regulator; CFTR, as well as membrane-bound receptors and ion channels associated with CFTR), glucose-galactose malabsorption syndrome (Na.sup.+ /glucose cotransporter), hypercholesterolemia (low-density lipoprotein (LDL) receptor), and forms of diabetes mellitus (insulin receptor), abnormal hormonal secretion linked to disorders including diabetes insipidus (vasopressin), hyper- and hypoglycemia (insulin, glucagon), Grave's disease and goiter (thyroid hormone), Cushing's and Addison's diseases (adrenocorticotrophic hormone; ACTH), disorders related to excessive secretion of biologically active peptides by tumor cells including fasting hypoglycemia due to increased insulin secretion from insulinoma-islet cell tumors, hypertension due to increased epinephrine and norepinephrine secreted from pheochromocytomas of the adrenal medulla and sympathetic paraganglia, carcinoid syndrome, which includes abdominal cramps, diarrhea, and valvular heart disease, caused by excessive amounts of vasoactive substances (serotonin, bradykinin, histamine, prostaglandins, and polypeptide hormones) secreted from intestinal tumors. Ectopic synthesis and secretion of biologically active peptides (peptides not expected from a tumor) includes ACTH and vasopressin in lung and pancreatic cancers; parathyroid hormone in lung and bladder cancers; calcitonin in lung and breast cancers; and thyroid-stimulating hormone in medullary thyroid carcinoma.

An association between the level of expression and/or activity of the present protein with the presence or absence of any condition associated with abnormal vesicular trafficking and/or secretion, such as any of the above-listed disorders, can readily be assessed by detecting the level of expression or activity of the protein by, e.g., Northern blot, western blot, ELISA, or any standard in

vitro or in vivo assay for protein activity, and correlating the observed level or expression or activity with the presence or absence of the disorder. For those disorders found to be positively associated with the protein of the invention, a diagnostic or screening assay can be readily developed where the detection of an elevated level of protein or protein activity is indicative of the presence of the disease, or of a propensity to develop the disease. Further, any such diseases or conditions can be treated or prevented by inhibiting the expression or activity of the protein, for example by administering to a patient suffering from the disorder any inhibitor including, but not limited to, antibodies, antisense oligonucleotides, dominant negative forms of the protein, and small molecule inhibitors of protein expression or activity. Alternatively, disorders negatively associated with the protein of the invention can be diagnosed or screened for by detecting the level of the present protein or protein activity, where a decreased level of the protein or protein activity is indicative of the presence of the disease, or of a propensity to develop the disease. Such disorders that are negatively associated with the protein of the invention can be treated or prevented by increasing the level of the protein or protein activity, for example by administering to a patient any of a number of agents including, but not limited to, the protein itself, a polynucleotide encoding the protein, or a heterologous compound that enhances the expression or activity of the protein.

Cancer cells secrete excessive amounts of hormones or other biologically active peptides. Therefore, in another embodiment, antagonists or inhibitors of the protein of the invention may be administered to a subject to treat or prevent cancers by inhibiting the traffic activity in transformed cells. Any type of cancer can be treated or prevented in this way, including, but not limited to, adenocarcinoma, sarcoma, melanoma, lymphoma, and leukemia. In preferred embodiments, the cancers include cancers of glands, tissues, and organs involved in secretion or absorption, such as prostate, pancreas, lung, tongue, brain, breast, bladder, adrenal gland, thyroid, liver, uterus, ovary, kidney, testes, and organs of the gastrointestinal tract including small intestine, colon, rectum, and stomach. In a particular aspect, antibodies which are specific for the protein of the invention may be used directly as an antagonist, or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express the protein of the invention. In addition, the elevated amount of the protein of the invention in tumor cells can readily be used to diagnose or screen for cancer, e.g. by measuring and comparing the level of the protein in a cell to that of a control cell using a specific antibody detected by FACS or using any other detection method known to those of skill in the art.

Protein of SEQ ID NO:297 (181-3-3-0-C9-CS)

The protein of SEQ ID No:297, encoded by the cDNA of SEQ ID NO:56, is homologous to synaptogyrin 1 (Trembl ID: Q9UGZ4). The protein of the invention is highly expressed in the brain and fetal brain, fetal liver and the testis.

The protein of SEQ ID No:297 is a splice variant of synaptogyrin 1. The splicing of the cDNA of SEQ ID NO:56 is different for exon 3: whereas exon 3 of synaptogyrin 1 is 238 base-pair long, exon 3 of SEQ ID NO:56 is 345 base-pair long. This introduces a frameshift and a stop codon. Thus, the protein of SEQ ID NO:297 is identical to synaptogyrin 1 up to and including
5 amino acid 122, the remaining 22 amino acids are entirely different. When compared to synaptogyrin 1, the protein of the invention presents the same N-terminal domain (which is highly conserved in all synaptogyrins) and 2 of the 4 transmembrane helices. Preferred polypeptides of the invention are those that comprise amino acids 1 to 16, which make up the N terminal cytoplasmic domain of the protein and which are highly conserved among all members of the synaptogyrin
10 family (Kedra D et al. – Hum Genet. – 1998, 103(2):131-141). Other preferred polypeptides of the invention are those that comprise amino acids 25 to 45 and/or 68 to 88, which make up the two transmembrane alpha helices. Thus it is believed that the protein of the invention is a member of the synaptogyrin family.

Synaptogyrins are closely related to proteins of the synaptophysin family, both of which are
15 involved in neurotransmission and more generally in exocytosis and vesicle trafficking. Members of the synaptogyrin family include synaptogyrin 1 (with splice variants 1a, 1b and 1c), cellugyrin (synaptogyrin 2) and synaptogyrin 3. This family of proteins is also evolutionarily conserved, as homologues to human synaptogyrin 1 have been found in rats, mice, and *C. elegans*. Synaptogyrins and synaptophysin are among the most abundant vesicle components--together they account for
20 more than 10% of the total vesicle membrane proteins. Although synaptogyrins do not appear to be required for exocytosis itself (apparently because synaptogyrins and synaptophysins have overlapping functions), they are essential for the normal regulation of exocytosis.

The normal function and organization of eukaryotic cells is dependent on the transport of various vesicles that selectively shuttle membrane and cargo between distinct compartments of the
25 secretory and endocytotic pathways. A number of key proteins involved in membrane targeting and exocytosis have been identified, and a fundamental set of interactions has been defined and placed into a model called the SNARE (Soluble N-ethylmaleimide-sensitive Attachment protein REceptor) hypothesis (Rothman J – Nature – 1994, 372: p55-63). According to the SNARE hypothesis, vesicles dock to a target membrane through the interaction of complementary sets of vesicular (v-
30 SNARE) and target (t-SNARE) membrane proteins. Our understanding of vesicle trafficking has, to a large extent, been facilitated by characterization of synaptic vesicles in neurons. In synaptic vesicle exocytosis, the vesicular protein synaptobrevin and synaptogyrin (also called Vesicle-Associated Membrane Protein; VAMP) are the v-SNARE, and the plasma membrane-associated protein SNAP-25 (Synaptosomal-Associated Protein of 25 kDa) and syntaxin 1 function as t-
35 SNARE. Formation of the SNARE complex (or core complex) is followed by recruitment of the cytosolic proteins alpha, beta and gamma SNAP (Soluble N-ethylmaleimide-sensitive Attachment Protein) and NSF (N-ethylmaleimide-Sensitive Factor), which are required for membrane fusion.

In transfected PC12 cells, synaptogyrin 1 and synaptophysin 1 are as effective as tetanus toxin light chain in inhibiting exocytosis (Sugita S. et al. – J Biol Chem. – 1999, 274(27):18893-901), suggesting that these proteins are strong regulators of exocytosis. More recently, synaptogyrins have been found to have an essential function in synaptic plasticity (Janz R. et al – Neuron. – 1999, 5 24(3):687-700).

The etiology of numerous human diseases and disorders can be attributed to defects in the trafficking of proteins to organelles or the cell surface. For example, defects in the trafficking of membrane-bound receptors and ion channels have been implicated in cystic fibrosis (cystic fibrosis transmembrane conductance regulator; CFTR), glucose-galactose malabsorption syndrome 10 (Na.sup.+ /glucose cotransporter), hypercholesterolemia (low-density lipoprotein (LDL) receptor), and forms of diabetes mellitus (insulin receptor). Abnormal hormonal secretion has been linked to disorders including diabetes insipidus (vasopressin), hyper- and hypoglycemia (insulin, glucagon), Grave's disease and goiter (thyroid hormone), and Cushing's and Addison's diseases (adrenocorticotrophic hormone; ACTH).

15 Further, cancer cells secrete excessive amounts of hormones or other biologically active peptides. Disorders related to excessive secretion of biologically active peptides by tumor cells include: fasting hypoglycemia due to increased insulin secretion from insulinoma-islet cell tumors; hypertension due to increased epinephrine and norepinephrine secreted from pheochromocytomas of the adrenal medulla and sympathetic paraganglia; and carcinoid syndrome, which includes 20 abdominal cramps, diarrhea, and valvular heart disease, caused by excessive amounts of vasoactive substances (serotonin, bradykinin, histamine, prostaglandins, and polypeptide hormones) secreted from intestinal tumors. Ectopic synthesis and secretion of biologically active peptides (peptides not expected from a tumor) includes ACTH and vasopressin in lung and pancreatic cancers; parathyroid hormone in lung and bladder cancers; calcitonin in lung and breast cancers; and thyroid-stimulating 25 hormone in medullary thyroid carcinoma.

In one embodiment, the invention relates to methods and compositions using the protein of the invention or part thereof as a new marker protein to selectively identify secretory and endocytic traffic, preferably in the endoplasmic reticulum and more preferably in the late Golgi cisternae. For example, the protein of the invention or part thereof may be detected using specific antibodies 30 generated against the protein using any technique known to those skilled in the art. Such organelle-specific antibodies may then be used to identify cells with disrupted trafficking systems such as in differentiated tumor cells or to differentiate specific organelle types in a cell cross-section using immunochemistry. In addition, the protein of the invention can be used to specifically identify cells of the brain, fetal brain, fetal liver and the testis, tissues in which the protein is overexpressed.

35 Another embodiment of the present invention relates to methods of targeting heterologous compounds, such as polypeptides or polynucleotides, to the components of the secretory machinery by recombinantly or chemically fusing a fragment of the protein of the invention to the heterologous

polypeptide or polynucleotide. Such fusion proteins may be engineered to contain a cleavage site located between a sequence encoding the protein of the invention and the heterologous protein sequence, so that the protein of the invention may be cleaved and purified away from the heterologous moiety. Such heterologous compounds may be targeted to the secretory pathway to modulate ER-Golgi endocytic and secretory activities. In one embodiment, the protein of the invention can be used to screen peptide libraries for inhibitors of traffic activity, as detected by the accumulation of ER membranes or Golgi vesicles as described in Conchon et al., supra.

In still another embodiment, the protein of the invention is used to diagnose, prevent and/or treat any of a number of disorders in which trafficking and/or the fusion machinery is affected, including, but not limited to, endocrine, secretory, inflammatory, and gastrointestinal disorders, such as cancer, cystic fibrosis (cystic fibrosis transmembrane conductance regulator; CFTR, as well as membrane-bound receptors and ion channels associated with CFTR), glucose-galactose malabsorption syndrome (Na.sup.+ /glucose cotransporter), hypercholesterolemia (low-density lipoprotein (LDL) receptor), and forms of diabetes mellitus (insulin receptor), abnormal hormonal secretion linked to disorders including diabetes insipidus (vasopressin), hyper- and hypoglycemia (insulin, glucagon), Grave's disease and goiter (thyroid hormone), Cushing's and Addison's diseases (adrenocorticotrophic hormone; ACTH), disorders related to excessive secretion of biologically active peptides by tumor cells including fasting hypoglycemia due to increased insulin secretion from insulinoma-islet cell tumors, hypertension due to increased epinephrine and norepinephrine secreted from pheochromocytomas of the adrenal medulla and sympathetic paraganglia, carcinoid syndrome, which includes abdominal cramps, diarrhea, and valvular heart disease, caused by excessive amounts of vasoactive substances (serotonin, bradykinin, histamine, prostaglandins, and polypeptide hormones) secreted from intestinal tumors. Ectopic synthesis and secretion of biologically active peptides (peptides not expected from a tumor) includes ACTH and vasopressin in lung and pancreatic cancers; parathyroid hormone in lung and bladder cancers; calcitonin in lung and breast cancers; and thyroid-stimulating hormone in medullary thyroid carcinoma.

An association between the level of expression and/or activity of the present protein with the presence or absence of any condition associated with abnormal vesicular trafficking and/or secretion, such as any of the above-listed disorders, can readily be assessed by detecting the level of expression or activity of the protein by, e.g., Northern blot, western blot, ELISA, or any standard in vitro or in vivo assay for protein activity, and correlating the observed level or expression or activity with the presence or absence of the disorder. For those disorders found to be positively associated with the protein of the invention, a diagnostic or screening assay can be readily developed where the detection of an elevated level of protein or protein activity is indicative of the presence of the disease, or of a propensity to develop the disease. Further, any such diseases or conditions can be treated or prevented by inhibiting the expression or activity of the protein, for example by administering to a patient suffering from the disorder any inhibitor including, but not limited to,

antibodies, antisense oligonucleotides, dominant negative forms of the protein, and small molecule inhibitors of protein expression or activity. Alternatively, disorders negatively associated with the protein of the invention can be diagnosed or screened for by detecting the level of the present protein or protein activity, where a decreased level of the protein or protein activity is indicative of the presence of the disease, or of a propensity to develop the disease. Such disorders that are negatively associated with the protein of the invention can be treated or prevented by increasing the level of the protein or protein activity, for example by administering to a patient any of a number of agents including, but not limited to, the protein itself, a polynucleotide encoding the protein, or a heterologous compound that enhances the expression or activity of the protein.

10 Cancer cells secrete excessive amounts of hormones or other biologically active peptides. Therefore, in another embodiment, antagonists, inhibitors, or other modulators of the protein of the invention may be administered to a subject to treat or prevent cancers by inhibiting the traffic activity in transformed cells. Any type of cancer can be treated or prevented in this way, including, but not limited to, adenocarcinoma, sarcoma, melanoma, lymphoma, and leukemia. In preferred
15 embodiments, the cancers include cancers of glands, tissues, and organs involved in secretion or absorption, such as prostate, pancreas, lung, tongue, brain, breast, bladder, adrenal gland, thyroid, liver, uterus, ovary, kidney, testes, and organs of the gastrointestinal tract including small intestine, colon, rectum, and stomach. In a particular aspect, antibodies which are specific for the protein of the invention may be used directly as an antagonist, or indirectly as a targeting or delivery
20 mechanism for bringing a pharmaceutical agent to cells or tissues which express the protein of the invention.

In addition, the present protein can be used to diagnose, treat, and prevent any neurological or psychiatric disorder or condition associated with abnormal neurotransmitter release, such as depression, which is associated with decreased serotonin secretion, or any neurological function,
25 e.g. memory, which could be enhanced or otherwise modulated by altering the quantity, frequency, or any other property of neurotransmitter release in one or more cell types in the nervous system.

Proteins of SEQ ID NOs:247 and 246 (internal designations 105-031-2-0-D3-CS and 105-031-1-0-A2-CS)

The protein of SEQ ID NOs:247 and 246, encoded by the cDNAs of SEQ ID NOs:6 and 5,
30 respectively, are overexpressed in liver, pancreas, and prostate. The proteins of the invention are strongly homologous to the human membrane-bound protein PRO836 (GENSEQP accession number: W63687), and to the human secreted protein 7 (GENSEQP accession number: Y57941). The proteins of the invention also share homology with the chaperone-associated protein, SLS1p, found in yeast *Yarrowia lipolytica* (GENPEPT accession number Z50154), having 27% identity
35 from amino-acids 68 to 340 of protein of SEQ ID No:247. In addition, the proteins of SEQ ID NOs:247 and 246 share homology with two Hsp70 family proteins, Hsp-binding protein 1 found in

mice (GENSEQP accession number: Z50154), and human species (GENPEPT accession number: AF093420), and Hsp-binding protein 2 found in human species (GENPEPT accession number: AF187859).

The proteins of the invention are related to a yeast lumen protein of the endoplasmic reticulum, SLS1p. This protein acts in the preprotein translocation process, interacting directly with translocating polypeptides to facilitate their transfer and/or help their folding in the endoplasmic reticulum (Boisrame et al. J Biol Chem 1996; 271:11668-75). In addition, Sls1p is believed to act as a cofactor of the chaperon protein Kar2 (Boisrame et al. J Biol Chem 1998; 273:30903-8 ; Kabani et al. Gene 2000; 241:309-15). Thus, the proteins of the invention are presumed to have similar cellular functions as those of chaperones. Such functions include a number of cellular processes, such as protein folding, disassembly of oligomeric protein structures, regulation of apoptosis, protein degradation, protein translocation in the endoplasmic reticulum, and antigen-presentation (Bukau et al. Cell 1998; 92:351-66). Chaperones are also involved in a number of disorders, especially autoimmune diseases such as type 1 diabetes, rheumatoid arthritis, systemic lupus erythematosus, Sjogren syndrome, and mixed connective tissue disease (Feige et al. EXS 1996; 77:359-73; Feili-Hariri et al. J Autoimmun 2000; 14:133-42). Chaperones are also involved in various disorders including tuberculosis and leprosy (Zugel et al. Clin Microbiol Rev 1999; 12:19-39), neurogenerative disorders such as Alzheimer and Parkinson diseases (Yoo et al. J Neural Transm Suppl 1999; 57:315-22), and malignant disorders (Csermely et al. Pharmacol Ther 1998; 79:129-68). In addition, a growing body of evidence suggests the involvement of the Hsp60 chaperone in the development of atherosclerosis (Xu et al. Circulation 2000; 102:14-20). Thus, the present proteins, which are presumed to be co-factors of a chaperon as summarized above, are believed to have analogous cellular functions and to be involved in similar pathological processes.

In one embodiment, the present invention provide methods of using the present proteins to identify specific cell types in vitro and in vivo. For example, as chaperone proteins are often upregulated in response to cellular stress, the detection of cells expressing elevated levels of the proteins provides a tool for detecting cells under stress. As cellular stress has been implicated in a number of disorders, such as cardiovascular disorders, neurodegenerative disorders, and cancer, the ability to detect such stress thus provides a diagnostic or screening tool for such conditions. In addition, the present polypeptides and polynucleotides can be used to identify liver, pancreas, and prostate tissues, and cells derived from these tissues. The ability to specifically visualize such tissues and cells is useful for a number of applications, including to determine the origin or identity of, e.g. cancerous cells, as well as to facilitate the identification of particular cells and tissues for, e.g. the evaluation of histological slides.

In addition, the present polypeptides and polynucleotides can be used to develop diagnostic and screening assays for diseases characterized by an abnormal level or activity of the protein of SEQ ID NOs:247 and 246. Such disorders include, but are not limited to, infectious diseases,

neurogenerative disorders such as Alzheimer's and Parkinson's diseases, schizophrenia, alopecia, aging, atherosclerosis, malignant disorders of various types, and autoimmune diseases including type 1 diabetes, rheumatoid arthritis, systemic lupus erythematosus, Sjogren syndrome, and mixed connective tissue disease. Such assays can be performed using any biological sample, such as
5 serum or plasma.

In still another embodiment, the proteins of the invention or part thereof can be used to prevent cells from undergoing apoptosis. Specifically, as chaperone proteins have been shown to protect cells from apoptosis, any method of increasing the level or activity of the present protein can be used to prevent cells from undergoing apoptosis, in vitro or in vivo. For example, a
10 polynucleotide encoding a protein of SEQ ID NO:247 or 246, or any fragment or derivative thereof, can be introduced into cells, e.g. in a vector, wherein the protein is expressed in the cells. Alternatively, a protein of SEQ ID NO:247 or 246 itself can be administered to cells, preferably in a formulation that leads to the internalization of the protein by the cells. Also, any compound that increases the expression or activation of the proteins within the cells can be administered.
15 Preventing cells from undergoing apoptosis can be used for any of a large number of purposes, including, but not limited to, to prevent the death of cells being grown in culture, to prevent in a patient the apoptosis associated with any of a number of disorders, or to prevent apoptosis in cells of a patient undergoing a treatment that increases the level of cellular stress, such as chemotherapy.

In another embodiment, inhibiting the proteins of the invention can be used to induce
20 apoptosis in undesired cells. Such inhibition can be accomplished in any of a number of ways, including, but not limited to, using antibodies, antisense sequences, dominant negative forms of the protein, or small molecule inhibitors of the expression or activity of the proteins. Such induction of apoptosis can be used to eliminate any undesired cells, for example cancer cells, in a patient. Preferably, such inhibitors are targeted specifically to the undesired cells in the patient.

25 In another embodiment, various disorders can be treated, attenuated and/or prevented by a protein of SEQ ID NOs:247 or 246, or part thereof, or any other compound that can affect the level or activity of the proteins such as nucleic acids, antibodies, or chemical substances. In a preferred embodiment, proteins or other compounds directed to the proteins of the invention can be used to treat or prevent disorders in which the activity or level of the proteins of SEQ ID NO:247 or 246 is
30 unbalanced. Such diseases include, but are not limited to, infectious diseases, neurogenerative disorders as Alzheimer and Parkinson diseases, schizophrenia, alopecia, aging, atherosclerosis, malignant disorders of various types, and autoimmune diseases including type 1 diabetes, rheumatoid arthritis, systemic lupus erythematosus, Sjogren syndrome, mixed connective tissue disease, malignant disorders, autoimmune and any other neurodegenerative disorder. In another
35 embodiment, the proteins of SEQ ID NO:247 or 246 or part thereof can be used as vaccines for various disorders including, but not limited, to cancer (Wang et al. Immunol Invest 2000;29:131-7),

tuberculosis (Silva et al. *Microbes Infect* 1999;1:429-35), diabetes (*Int Immunol* 1999;11:957-66), and atherosclerosis (Xu et al. *Arterioscler Thromb* 1992;12:789-99).

Protein of SEQ ID NO:389 (internal designation 109-003-1-0-G4-CS)

The protein of SEQ ID NO:389 is encoded by the cDNA of SEQ ID NO:148. Accordingly,
5 it will be appreciated that all characteristics and uses of the polypeptide of SEQ ID NO:389 described throughout the present application also pertain to the polypeptide encoded by the human cDNA of clone 109-003-1-0-G4-CS. In addition, it will be appreciated that all characteristics and uses of the nucleic acid of SEQ ID NO:148 described throughout the present application also pertain to the human cDNA of clone 109-003-1-0-G4-CS. The protein of SEQ ID NO:389 is highly
10 homologous to two human proteins encoded by genes listed in Genbank under accession numbers AF143723 and AF112210, the disclosures of which are incorporated herein by reference in their entireties.

The polypeptide encoded by Genbank accession numbers AF143723 and AF112210 belong to the Hsp70 protein family (even though one of them has erroneously been attributed to the related
15 Hsp60 family). Many genes encoding "Hsps" (heat shock proteins) have been cloned and sequenced, including, for example, human hsp70 (GenBank Accession Nos. M11717 and M15432; see also Hunt and Morimoto, 1985, *Proc. Natl. Acad. Sci. USA* 82: 6455-6459, the disclosures of which are incorporated herein by reference in their entireties), human hsp90 (GenBank Accession No. X15183; see also Yamazaki et al., 1989, *Nucleic Acids Res.* 17: 7108, the disclosures of which
20 are incorporated herein by reference in their entireties), and human gp96 (GenBank Accession No. M33716; see also Maki et al., 1990, *Proc. Natl. Acad. Sci. USA* 87: 5658-5662, the disclosures of which are incorporated herein by reference in their entireties).

The protein of SEQ ID NO:389 and the two homologs mentioned above are actually closer to yeast members of the family than to the human Hsp70, which makes the corresponding genes
25 previously unidentified human members of the family. Both the Pfam and Prosite Hsp70 signatures (respectively the "HSP70" Pfam model from amino acid position 3 to 509 and the PS01036 Prosite motif from position 332 to 346) are recognized within the protein of SEQ ID NO: 389. The protein of SEQ ID NO:389 differs from the protein encoded by AF112210 at amino-acid positions 282, 312 and 326, and from the protein encoded by AF143723 at amino acid position 15 and 326.

30 Heat shock proteins are a family of molecular chaperone proteins which have long been known to play essential roles in a multitude of intra-and intercellular processes, including protein synthesis and folding, vesicular trafficking, and antigen processing and presentation. Hsps are among the most highly conserved proteins known, and carry out many of their regulatory activities via protein-protein interactions. Historically they were identified by induction under conditions of
35 stress, during which they are now known to provide an essential action of preventing aggregation

and assisting refolding of misfolded proteins. The major stress proteins accumulate to very high levels in stressed cells but occur at low to moderate levels in cells that have not been stressed.

Hsp70 is one member of the heat shock protein family. (Milner, C. M. and Campbell, R. D. Immunogenetics 32: 242-251 (1990); Genbank Accession No. M59828, the disclosures of which are incorporated herein by reference in their entirety). The 70kD heat shock protein is a highly conserved, ubiquitous protein involved in chaperoning proteins to various cellular organelles. Contrary to other members of the Hsp family, it is highly inducible in mammals. Although Hsp70 is barely detectable at normal temperatures, it becomes one of the most actively synthesized proteins in the cell upon heat shock (Welch et al., 1985, J. Cell. Biol. 101:1198-1211, the disclosure of which is incorporated herein by reference in its entirety). In contrast, the Hsp90 and Hsp60 proteins are abundant at normal temperatures in most, but not all, mammalian cells and are further induced by heat (Lai et al., 1984, Mol. Cell. Biol. 4:2802-10; van Bergen en Henegouwen et al., 1987, Genes Dev., 1:525-31, the disclosures of which are incorporated herein by reference in their entirety). Furthermore the Hsp70 proteins act as monomers whereas the functionally related Hsp60 proteins are associated in vivo within large double ring assemblies of nearly a million daltons. The various actions of the Hsps all rely basically on their ability to complex polypeptide segments, preferably hydrophobic, and to stabilize them in an extended conformation in an ATP-dependent manner. The complexed polypeptides can be antigenic peptides (in which case the Hsps help directing them to the major histocompatibility complexes for presentation) or misfolded proteins which are facilitated to adopt the proper conformation by repeated cycles of binding to Hsps followed by release/refolding (see Bukau, B. and Horwich L., 1998, Cell 92: 351-366, the disclosure of which is incorporated herein by reference).

On the basis of the above information, it is believed that the protein of SEQ ID NO:389 is a member of the human Hsp70 family. Accordingly, the protein of SEQ ID NO:389 may play a role in protein synthesis/folding, cellular trafficking, antigen processing, the cellular stress response and the immune response in immuno-competent cell types. Additional information regarding the protein of SEQ ID NO:389 may be obtained by performing a binding assay with a consensus Hsp70 substrate using the methods described in Rüdiger et al., 1997, EMBO J. 16, 1501-1507, the disclosure of which is incorporated herein by reference in its entirety.

One embodiment of the present invention relates to methods of using the protein of SEQ ID NO:389 or fragments comprising at least 5, 8, 10, 12, 15, 20, 25, 30, 35, 40, 50, 60, 75, 100, 150, or 200 consecutive amino acids thereof, or fragments having a desired biological activity as a stabilizing adjuvant to slow down protein degradation, boost the yields of recombinant proteins or regenerate denatured proteins. In such an embodiment, the protein of SEQ ID NO:389 of fragment thereof is mixed with a composition comprising the protein for which it is desired to slow down degradation, boost yield, or regenerate denatured proteins under conditions which facilitate the desired result.

For example, numerous commercial assay kits commonly used by those skilled in the arts of molecular biology and biochemistry depend on the biological properties of proteins (mostly enzymes) which can be very short-lived in vitro due to the low stability of those proteins. An example is described in Eur. Patent DE4124286, the disclosure of which is incorporated herein by reference in its entirety, wherein the low intrinsic stability of test solutions used in optical tests is increased by addition of chaperone proteins, thus making the test more sensitive.

The protein of SEQ ID NO:389 may also be used to increase the yield or activity of recombinant proteins. In recombinant DNA technology, a major unsolved problem is the solubility and biological activity of the recombinantly overexpressed protein in a host, especially a bacterial or yeast host. Many eukaryotic proteins, especially the secreted ones, require for correct folding a specific cellular machinery which is lacking in bacterial hosts such as *E. coli* or becomes insufficient in mammalian/yeast cells due to high expression of the protein. The ability of the protein of SEQ ID NO:389 or fragments thereof to ensure proper folding of recombinant proteins may be utilized as follows. The protein of SEQ ID NO:389, may be coexpressed with the recombinant protein in bacterial or eukaryotic hosts to cause the hosts to express the heterologous proteins or polypeptides in a form having increased solubility and/or biological activity. For example, the protein of SEQ ID NO:389 or fragments thereof may be used in the methods described in PCT application WO 93/25681, the disclosure of which is incorporated herein by reference in its entirety. Alternatively the protein of SEQ ID NO:389 or fragments thereof may be exogenously added to the cell cultures as described in PCT application WO 00/08135, the disclosure of which is incorporated herein by reference in its entirety. Indeed PCT application WO 00/31113, the disclosure of which is incorporated herein by reference in its entirety, shows that when added exogenously to cells, Hsp70 is readily imported into both cytoplasmic and nuclear compartments. Preparation and purification of the protein of SEQ ID NO:389 or fragments thereof may be carried out as described in Patent US-6,007,821, the disclosure of which is incorporated herein by reference in its entirety.

The protein of SEQ ID NO:389 or fragments thereof, may further be used to regenerate denatured proteins. Recombinantly expressed proteins with poor biological activity are routinely denatured with a potent denaturing agent, such as guanidine hydrochloride, followed by refolding by dilution with a large amount of a diluent to reduce the concentration of the denaturing agent. However, this method often results in a poor refolding rate which may be significantly increased by addition of a cocktail of chaperone proteins in a fashion similar to that described for Hsp60 in Eur. Patent EP0650975, the disclosure of which is incorporated herein by reference in its entirety. The advantage of using a cocktail of chaperone proteins is to accommodate differences in binding specificity of the Hsp different families and the different members within each family. For instance, vertebrate actin is efficiently folded by the chaperonin of the eukaryotic cytosol (Gao et al., 1992, Cell 69:1043-1050, the disclosure of which is incorporated herein by reference in its entirety) but

not at all by Hsp60 (Tian et al., 1995, Nature 375:250-253, the disclosure of which is incorporated herein by reference in its entirety).

Another embodiment of the present invention relates to the use of the protein of SEQ ID NO:389 or fragments thereof to deliver heterologous compounds (proteins, peptides, or DNA) to specific cellular compartments, preferably the cytoplasm and the nucleus. If desired, the protein of SEQ ID NO:389 or a fragment thereof may be fused to the heterologous compound. For example, the protein of SEQ ID NO:389 or fragments thereof may be used to chaperone compounds into cells using the methods described in PCT application WO 00/31113, the disclosure of which is incorporated herein by reference in its entirety. In the methods described in WO 00/31113, Hsp70 was used to deliver NF-KB, a key transcriptional regulator of inflammatory responses, into the nuclear compartment. It was shown that a fusion protein composed of a Cterminal Hsp70 peptide and amino acids 37-409 of the p50 subunit of NF-KB was directed into the nucleus of cells, could bind DNA specifically, and activated kappa Ig expression and TNFa production.

In one embodiment of the present invention, the protein of SEQ ID NO:389 or a fragment thereof may be used in human therapy as a modulator of immune response. Disease states which may be treated by Hsp70, fragments thereof, and/or Hsp70 complexes of the present invention include transplant rejection (see US5,891,653, the disclosure of which is incorporated herein by reference in its entirety) and autoimmune diseases, such as insulin dependent diabetes mellitus, rheumatoid arthritis, multiple sclerosis, juvenile diabetes, asthma, and inflammatory bowel disease, as well as inflammatory diseases, cancer, viral replication diseases and vascular diseases as described in the following patents, each of which is incorporated herein by reference in its entirety: US6,007,821; WO 00/31113; WO 99/18801 (treatment of auto-immune diseases), US6,017,540; US6,017,544; AU3425899; WO 99/54464; US5,837,251; US5,830,464; WO 98/34642; WO 98/34641; US5,750,119; WO 97/10001; WO 96/10411(cancer treatment); DE19813760, DE19813759 (both autoimmune disease and cancer).

The protein of SEQ ID NO:389 or fragments thereof may also be used to treat or ameliorate autoimmune disease. In this embodiment, compositions of complexes of heat shock/stress proteins (including, but not limited to the protein of SEQ ID NO:389) are administered to an individual suffering from an autoimmune disease. The complexes may be comprised of the protein of SEQ ID NO:389 or fragments thereof alone or may include other heat shock/stress proteins. In one embodiment, the protein of SEQ ID NO:389 or a fragment thereof is bound noncovalently to antigenic molecules and administered to individuals suffering from autoimmune disease to suppress the autoimmune response. Alternatively, compositions comprising the protein of SEQ ID NO:389 or fragments thereof in an un-complexed form (i.e., free of antigenic molecules) may also be administered to an individual suffering from autoimmune disease to suppress the immune response (see Patent US6,007,821, the disclosure of which is incorporated herein by reference in its entirety).

The ability of stress proteins to chaperone the antigenic peptides of the cells from which they are derived allows them to be used to isolate the antigenic peptides expressed in a tumor. In this embodiment of the present invention, complexes comprising the protein of SEQ ID NO:389 or fragments thereof and an antigenic peptide expressed by the tumor are isolated. The isolated
5 complexes are administered back to the individual from which they were obtained in order to elicit an immune response against the tumor. Accordingly, this approach circumvents the necessity of isolating and characterizing specific tumor antigens and enables the skilled artisan to readily prepare immunogenic compositions effective against a tumor in an individual (see Patent US6,017,544, the disclosure of which is incorporated herein by reference in its entirety).

10 The protein of SEQ ID NO:389 may also be used to diagnose bladder cancer. The segment of the protein of SEQ ID NO:389 extending between amino acid positions 1 through 187 is more than 99% identical to a polypeptide which is linked to bladder cancer. (See Eur. Patent DE19818620, the disclosure of which is incorporated herein by reference in its entirety). The 187
amino-acid long polypeptide described in DE19818620 was identified as the partial product of the
15 only gene for which expression was significantly altered in a bladder tumour compared to a healthy bladder. In another embodiment of the present invention, the protein of SEQ ID NO:389 or a fragment thereof thereof may be used to diagnose disorders associated with altered intercellular communication or secretion. In such techniques, the level of the protein of SEQ ID NO:389 in an individual is measured using techniques such as those described herein. The level of the protein of
20 SEQ ID NO:389 in the individual is compared to the level in normal individuals. An altered level of the protein of SEQ ID NO:389 relative to normal individuals suggests that the individual is suffering from bladder cancer. The level of the protein of SEQ ID NO:389 present in the individual may determined by contacting a sample from the individual with an antibody directed against the polypeptide of SEQ ID NO:389. Alternatively, the level of the protein of SEQ ID NO:389 in the
25 individual may be measured by determining the level of RNA encoding the protein of SEQ ID NO:389 in the sample. RNA levels may be measured using nucleic acid arrays or using techniques such as in situ hybridization, Northern blots, dot blots or other techniques familiar to those skilled in the art. If desired, an amplification reaction, such as a PCR reaction, may be performed on the nucleic acid sample prior to analysis. The level of RNA in the sample is compared to RNA levels
30 in normal individuals to determine whether the individual is suffering from bladder cancer.

Antibodies against the protein of the protein of SEQ ID NO:389 or nucleic acid probes complementary to the sequence encoding the protein of SEQ ID NO:389 may also be used as a prognosis of tumor recurrence in breast as described in Patent US Patent No.: 5,188,964, the disclosure of which is incorporated herein by reference in its entirety. As described in U.S. Patent
35 No. 5,188,964, specific levels of the stress response proteins (including Hsp70) were identified, above which the probability of tumor recurrence is highly significant. Accordingly, the levels of the protein of SEQ ID NO:389 or RNA encoding the protein of SEQ ID NO:389 may be determined

from in a sample from an individual who has experienced a breast tumor in the past. Protein or RNA levels may be measured as described herein. If the protein or RNA levels exceed the levels above which tumor occurrence is likely, an appropriate course of treatment may be initiated.

In another embodiment of the present invention, the protein of SEQ ID NO:389 may be used to promote tissue repair and/or increase cell survival in stress conditions such as hypoxia, oxidative stress, genotoxic agents and more generally harmful conditions leading to programmed cell death. The beneficial effect is produced either by protecting the cell proteins from premature denaturation/degradation or by directly inhibiting a signal transduction pathway leading to programmed cell death (Gabai VL. et al., 1998, FEBS Lett. 438:1-4, the disclosure of which is incorporated herein by reference in its entirety). Those conditions include but are not limited to infarction, heart surgery, stroke, neurodegenerative diseases, epilepsy, trauma, atherosclerosis, restenosis after angioplasty, and nerve damage. For example, it is known that hypoxic stress is a signal that increases the amount of Hsp70 in cardiac tissue, whereupon Hsp70 helps cells survive by binding to partially denatured proteins and assisting in the refolding of these proteins into more stable native structures. Such assistance would be extremely important in providing protection to the heart during periods of hypoxia such as during an infarct or during surgery when blood flow to the heart may be temporarily halted. Several groups have also shown that overproduction of Hsp70 leads to protection in several different models of nervous system injury (reviewed in Midori AY et al., 1999, Mol. Med. Today, 5:525-31, the disclosure of which is incorporated herein by reference in its entirety). Therapeutic methods for administering the protein of SEQ ID NO:389 or a fragment thereof include but are not limited to those disclosed in Patent WO 00/23093, the disclosure of which is incorporated herein by reference in its entirety.

Accordingly, it may be desirable to increase or decrease the level of the protein of SEQ ID NO:389 in an individual having a condition resulting from an increased or decreased level of the protein. In such embodiments, the protein of SEQ ID NO:389, or a fragment thereof, is administered to an individual in whom it is desired to increase or decrease any of the foregoing activities. The protein of SEQ ID NO:389 or fragment thereof may be administered directly to the individual or, alternatively, a nucleic acid encoding the protein of SEQ ID NO:389 or a fragment thereof may be administered to the individual. Alternatively, an agent which increases the activity of the protein of SEQ ID NO:389 may be administered to the individual. Such agents may be identified by contacting the protein of SEQ ID NO:389 or a cell or preparation containing the protein of SEQ ID NO:389 with a test agent and assaying whether the test agent increases the activity of the protein. For example, the test agent may be a chemical compound or a polypeptide or peptide.

Alternatively, the activity of the protein of SEQ ID NO:389 may be decreased by administering an agent which interferes with such activity to an individual. Agents which interfere with the activity of the protein of SEQ ID NO:389 may be identified by contacting the protein of

SEQ ID NO:389 or a cell or preparation containing the protein of SEQ ID NO:389 with a test agent and assaying whether the test agent decreases the activity of the protein. For example, the agent may be a chemical compound, a polypeptide or peptide, an antibody, or a nucleic acid such as an antisense nucleic acid or a triple helix-forming nucleic acid.

5 Protein of SEQ ID NO:250 (internal designation 105-053-4-0-E8-CS)

The protein of SEQ ID NO:250 is encoded by the cDNA of SEQ ID NO:9. It will be appreciated that all characteristics and uses of the polypeptide of SEQ ID NO:250 described throughout the present application also pertain to the polypeptide encoded by the human cDNA of clone 105-053-4-0-E8-CS. In addition, it will be appreciated that all characteristics and uses of the
10 nucleic acid of SEQ ID NO:9 described throughout the present application also pertain to the human cDNA of clone 105-053-4-0-E8-CS. The protein of SEQ ID NO:250 is found in prostate and exhibits extensive homologies to stretches of pancreatic zymogen granule membrane protein GP2 (Glycoprotein-2). In particular, the protein of SEQ ID NO:250 exhibits homologies to the GP2 proteins of human (SWISS-PROT accession number P55259, the disclosure of which is
15 incorporated herein by reference in its entirety), rat (SWISS-PROT accession number P19218, the disclosure of which is incorporated herein by reference in its entirety) and dogs (SWISS-PROT accession number P25291, the disclosure of which is incorporated herein by reference in its entirety). In fact, the amino acid sequence of SEQ ID NO:250 is completely identical to those of human GP2 sequences except that the protein of SEQ ID NO:250 is missing amino acids 62 to 484
20 from the human GP2 sequence. The protein of SEQ ID NO:250 contains two hydrophobic regions, namely the N-terminal signal peptide (amino acid residues 8-28) and the C-terminal transmembrane domain (amino acid residues 91-111).

GP2 (Glycoprotein-2) is the major membrane glycoprotein of secretory zymogen granule (ZG) membranes within pancreatic acinar cells (Fukuoka et al. 1990 Nuc. Acids Res., 18:5900;
25 Fukuoka et al. 1991 Proc. Natl. Acad. Sci., USA, 88:2898-2902; Fukuoka et al. 1992 Proc. Natl. Acad. Sci. USA, 89:1189-1193; Freedman, et al. 1993 Eur. J. Cell Biol. 61:229-238; Scheele et al. 1993 Pancreas :139-149; Freedman et al. 1994 Annals N.Y. Acad. Sci. 713:199-206, the disclosures of which are incorporated herein by reference in their entireties). GP2 homologues are also widely distributed among diverse epithelial tissues known to possess regulated secretory processes,
30 including parotid, submandibular gland, stomach, liver and lung (Fukuoka et al. 1992 Proc. Natl. Acad. Sci. USA, 89:1189-1193).

In addition to ZG membranes, GP2 is also located in pancreatic acinar cells in rough endoplasmic reticulum, Golgi, trans-Golgi components, condensing vacuoles, apical plasma membranes (APM), basolateral plasma membranes (BPM), and within ZGs and acinar lumina
35 (Scheele et al., 1994 Pancreas 9:139-149). GP2 is linked to the membrane of the ZG via a glycosylphosphatidyl inositol-anchor (GPI-anchor) (Fukuoka et al. 1991 Proc. Natl. Acad. Sci.

USA, 88:2898-2902; Lebel and Beattie 1988 Biochem. Biophys. Res. Comm. 254:1189-93, the disclosures of which are incorporated herein by reference in their entireties) and forms complexes, usually tetrameric complexes, below a pH of about 6.5.

During assembly of secretory granules within the trans-Golgi network (TGN), the low pH
5 of the TGN causes formation of GP2 complexes. These complexes bind to proteoglycans (PG), forming a fibrillar GP2/PG meshwork on the luminal surface of the ZG. The GP2/PG matrix may function in membrane sorting within the TGN, assembly of ZG membranes, inactivation of ZG membranes during granule storage, and regulation of ZG membrane trafficking at the apical plasma membrane. The GP2/PG matrix may also protect the luminal aspect of the granule membrane from
10 contact with secretory enzymes contained within the granules and facilitate the specific release of secretory enzymes during exocytosis at the apical plasma membrane.

The enzymes and the acidic milieu contained in the ZG are released into the lumen of the pancreas through exocytosis by acinar cells. The pH at the apical plasma membrane of the acinar cells, and of the pancreatic lumen in general, is maintained at an essentially neutral or alkaline pH
15 by the fluid and bicarbonate secreted by pancreatic ductal cells. The increased pH at the apical plasma membrane (relative to the acidic pH within the ZG) optimizes the conditions for enzymatic cleavage of the GPI anchor of GP2, resulting in release of GP2 and GP2/PG complexes from the apical membrane. (Scheele et al. (1994) Pancreas 9:139-149, the disclosure of which is incorporated herein by reference in its entirety). The form of GP2 produced by GPI-anchor cleavage is termed
20 globular GP2 (gGP2).

It is believed that the protein of SEQ ID NO:250 is a GP2 protein, and is thus likely involved in regulated membrane trafficking along apical secretory processes in a variety of epithelial cells.

Accordingly, the present invention includes the use of the protein of SEQ ID NO:250,
25 fragments comprising at least 5, 8, 10, 12, 15, 20, 25, 30, 35, 40, 50, 60, 75, 100, 150, or 200 consecutive amino acids thereof, or fragments having a desired biological activity in the modulation of membrane sorting within the trans-Golgi network, assembly of zymogen granule membranes, inactivation of zymogen granule membranes during granule storage, regulation of zymogen granule membrane trafficking at the apical plasma membrane, release of secretory enzymes during
30 exocytosis at the apical plasma membrane. In such embodiments, the protein of SEQ ID NO:250, or a fragment thereof, is administered to an individual in whom it is desired to increase or decrease any of the foregoing activities. The protein of SEQ ID NO:250 or fragment thereof may be administered directly to the individual or, alternatively, a nucleic acid encoding the protein of SEQ ID NO:250 or a fragment thereof may be administered to the individual. Alternatively, an agent
35 which increases the activity of the protein of SEQ ID NO:250 may be administered to the individual. Such agents may be identified by contacting the protein of SEQ ID NO:250 or a cell or preparation containing the protein of SEQ ID NO:250 with a test agent and assaying whether the

test agent increases the activity of the protein. For example, the test agent may be a chemical compound or a polypeptide or peptide.

Alternatively, the activity of the protein of SEQ ID NO:250 may be decreased by administering an agent which interferes with such activity to an individual. Agents which interfere with the activity of the protein of SEQ ID NO:250 may be identified by contacting the protein of SEQ ID NO:250 or a cell or preparation containing the protein of SEQ ID NO:250 with a test agent and assaying whether the test agent decreases the activity of the protein. For example, the agent may be a chemical compound, a polypeptide or peptide, an antibody, or a nucleic acid such as an antisense nucleic acid or a triple helix-forming nucleic acid.

10 In one embodiment, the invention relates to methods and compositions using the protein of the invention or part thereof as a marker protein to selectively identify tissues, preferably pancreas and prostate, or to distinguish between two or more possible sources of a tissue sample on the basis of the level of the protein of SEQ ID NO:250 in the sample. For example, the protein of SEQ ID NO:250 or fragments thereof may be used to generate antibodies using any techniques known to those skilled in the art, including those described therein. Such tissue-specific antibodies may then be used to identify tissues of unknown origin, for example, forensic samples, differentiated tumor tissue that has metastasized to foreign bodily sites, or to differentiate different tissue types in a tissue cross-section using immunochemistry. In such methods a tissue sample is contacted with the antibody, which may be detectably labeled, under conditions which facilitate antibody binding. The level of antibody binding to the test sample is measured and compared to the level of binding to control cells from pancreas or prostate or tissues other than pancreas or prostate to determine whether the test sample is from pancreas or prostate. Alternatively, the level of the protein of SEQ ID NO:250 in a test sample may be measured by determining the level of RNA encoding the protein of SEQ ID NO:250 in the test sample. RNA levels may be measured using nucleic acid arrays or using techniques such as in situ hybridization, Northern blots, dot blots or other techniques familiar to those skilled in the art. If desired, an amplification reaction, such as a PCR reaction, may be performed on the nucleic acid sample prior to analysis. The level of RNA in the test sample is compared to RNA levels in control cells from pancreas or prostate or tissues other than pancreas or prostate to determine whether the test sample is from pancreas or prostate.

30 In another embodiment, antibodies to the protein of the invention or part thereof may be used for detection, enrichment, or purification of membranes or zymogen granules using any techniques known to those skilled in the art. For example, an antibody against the protein of SEQ ID NO:250 or a fragment thereof may be fixed to a solid support, such as a chromatography matrix. A preparation containing membranes or zymogen granules is placed in contact with the antibody under conditions which facilitate binding to the antibody. The support is washed and then the membranes or zymogen granules are released from the support by contacting the support with agents which cause the membranes or zymogen granules to dissociate from the antibody.

In another embodiment of the present invention, the protein of SEQ ID NO:250 or a fragment thereof may be used to diagnose disorders associated with altered intercellular communication or secretion. In such techniques, the level of the protein of SEQ ID NO:250 in a patient is measured using techniques such as those described herein. The level of the protein of
5 SEQ ID NO:250 in the patient is compared to the level in control individuals. An elevated level or decreased level of the protein of SEQ ID NO:250 relative to control individuals suggests that the patient is suffering from a defect in intercellular communication or secretion.

In another embodiment, the protein of SEQ ID NO:250 or a fragment thereof is used to facilitate or decrease exocytosis. For example, the protein of SEQ ID NO:250 or fragment thereof
10 may be used to increase or decrease the release of secretory enzymes within pancreatic acinar cells or prostatic cells. Accordingly, the protein of the invention or part thereof may be used to diagnose, treat and/or prevent disorders associated with abnormal membrane trafficking including but not limited to viral or other infections, traumatic tissue damage, and hereditary diseases such as pancreatitis or prostatitis, invasive carcinomas and lymphomas. In such methods, the protein of
15 SEQ ID NO:250, a fragment of the protein of SEQ ID NO:250, or an agent which increases or decreases the activity of the protein of SEQ ID NO:250 is administered to an individual using techniques such as those described herein.

In another embodiment, the invention relates to methods of using the protein of SEQ ID NO:250 or a fragment thereof in the diagnosis of pancreatitis or prostatitis by detecting an elevation
20 in the level of the protein of SEQ ID NO:250, in a sample of bodily fluid, such as human blood, serum, or urine. The protein may be detected using any method known to those skilled in the art, including those described herein. In some embodiments, the protein of SEQ ID NO:250 or fragment thereof may be detected using the methods described in U.S. Patent Nos. 5436169 or 5663315, the disclosures of which are incorporated herein by reference in their entireties.

25 References :

- U.S. Patent Nos. 5,436,169; 5,663,315
- Nucleic Acids Research 18(9):5900, (1990)
- Proc. Natl. Acad. Sci. USA 88(7):2898-2902 (1991)
- Proc. Natl. Acad. Sci. USA 89:1189-1193 (1992)
- 30 Eur. J. Cell Biol. 61:229-238 (1993)
- Freedman et al., Annals N.Y. Acad. Sci. 713:199-206, 1994.
- Scheele et al., Pancreas 9(2):139-149, 1994.

Protein of Seq Id No: 274 (internal designation: 145-56-3-0-D5-CS)

The protein of SEQ ID No: 274 encoded by the cDNA of SEQ ID No: 33 is homologous to
35 the human RNA 3'-terminal phosphate cyclase-like protein 1 (*Rcl1*) (trEMBL accession number CAB89811) which is abundant in the nucleolus.

The RNA 3'-terminal phosphate cyclase, an enzyme originally identified in extracts from human HeLa cells and *Xenopus* oocyte nuclei, catalyzes the ATP-dependent conversion of the 3'-terminal phosphate group into a 2',3'-cyclic phosphodiester at the 3'-end of RNA, resulting in the

activation of the 3' end of RNA molecules. Database searches showed that genes encoding proteins similar to human and *E.coli* human RNA 3'-terminal phosphate cyclase are conserved among eukarya, bacteria and archaea, arguing for an essential function of the enzyme in RNA metabolism (Genschik P. et al. – EMBO J – 1998, 16, p.2955-2967). Similarly analysis of the human RNA 3'-terminal phosphate cyclases and related proteins from other organisms, indicated that they can be divided into 2 subfamilies referred to as RNA 3'-terminal phosphate cyclases (*Rtc*) and RNA 3'-terminal phosphate cyclase-like protein (*Rcl*). These 2 subfamilies share several sequence elements, including a nearly universally conserved amino acid sequence RGxxPxGGGx@ (where x stands for any, and @ for hydrophobic amino acids), designated originally as the cyclase signature, which corresponds to the Prosite signature, although structurally slightly different these 2 subfamilies of proteins have the same function and are involved in RNA metabolism. The cyclase signature is present in the protein of the invention (positions 157 to 167). In addition, this protein also displays other characteristic signatures of RNA 3'-terminal phosphate cyclase proteins (pfam signature from positions 1 to 368 and eMotif signatures from positions 12 to 44 and from positions 157 to 168).

3'-terminal phosphate cyclases (*Rtc* and *Rcl*) catalyze the conversion of 3'-terminal phosphate to a 2',3'-cyclic phosphodiester in a reaction dependent on ATP, other nucleoside triphosphates being much less active co-factors. With both enzymes, the cyclization of the 3'-phosphate at the 3'-end of RNA occurs by a three-step mechanism as follows : (a) adenylation of the enzyme by ATP; (b) the enzyme acts on RNA-N3' P to produce RNA-N3'PP5'A; (c) a non catalytic nucleophilic attack by the adjacent 2'hydroxyl on the phosphorus in the diester linkage to produce the cyclic end product.

RNA 3'-terminal phosphate cyclase proteins are involved in RNA processing. It has been demonstrated that several eukaryotic and prokaryotic RNA ligases require 2',3'-cyclic phosphate RNA ends which suggests that the enzyme is involved in generation or maintenance of cyclic termini in RNA ligation substrates, known to be required by several RNA ligases in both eukaryotes and prokaryotes. These ligases include 2 tRNA-splicing ligases, and the prokaryotic RNA ligase of unknown function that joins RNA ends via atypical 2',5'-phosphodiester (Arn E. et al. – RNA structure and Function - Cold spring Harbor Laboratory Press – 1998 p.695-726). The involvement of these ligases in nuclear pre-tRNA splicing is well documented (Zillmann et al. – Mol Cell Biol – 1991, 11, p5410-5416)(Phizicky E. et al. – J Biol Chem, 1992, 267, p4577-4582) but these enzymes might also function in the ligation of virusoids and viroids (Branch A. et al. – Science – 1982, 217, p1147-1149) (Kibertis et al. – EMBO J – 1985, 4, p817-827)

Alternatively, the cyclase could be responsible for producing cyclic phosphate 3'-ends identified in the spliceosomal U6 small nuclear RNA and some other small RNAs. Furthermore in yeast *Rcl* is associated to U3 small nucleolar RNP (U3 snoRNP) a central component of the 18S ribosomal RNA (rRNA) processing machinery in yeast and vertebrates (Billy E. et al. – EMBO J –

2000, 19, p2115-2126). However it seems that *Rcl* are not a structural component of U3 snoRNP and its association with U3 snoRNP occurs, most probably, in large macromolecular complexes representing nascent ribosomes. In yeast, depletion or inactivation of *Rcl* causes a defect in 18S mRNA synthesis, which leads to a decreased levels of 40S ribosomal sub-units, resulting in an accumulation of free 60S ribosomes and a fall in the amount of polysomes. In Yeast 18S, 5.8S and 25S rRNAs are derived from a long 35S precursor. This 35S pre-rRNA is normally cleaved at the A0 site, yielding 33S pre-rRNA. 33S rRNA is then processed rapidly at sites A1 and A2 to generate 20S pre-rRNA, which is further processed into mature 18S rRNA. Deletion or inactivation of *Rcl* leads to inhibition of processing at sites A0, A1 and A2 (Billy E. et al. – EMBO J – 2000, 19, p2115-2126).

It is believed that the protein of SEQ ID No: 274 or part thereof is involved in RNA processing, probably as a RNA 3'-terminal phosphate cyclase. Preferred polypeptides of the invention are polypeptides comprising amino acids 157 to 167, 1 to 368, 12 to 44 and 157 to 168. Other preferred polypeptides of the invention are fragments of SEQ ID No: 274 having any of the biological activities described herein. Assays of cyclase activity can be carried out using the Norit method as described in the article by Filipowicz (Filipowicz W et al.- Methods Enzymol. – 1990, 181, p.499-510), which disclosure is hereby incorporated by reference in its entity, or any other techniques known to those skilled in the art.

Thus, an embodiment of the present invention relates to compositions and methods of using the protein of the invention or part thereof in in vitro RNA manipulation to isolate small nucleolar RNPs especially, but not limited to U3 snoRNP from biological samples, using immunoprecipitation techniques (Billy E. et al. – EMBO J – 2000, 19, p2115-2126), which disclosure is hereby incorporated by reference in its entity, or any other techniques known to those skilled in the art.

In another embodiment, the protein of the invention or part thereof is used to develop antagonists of the protein of the invention or part thereof in order to inhibit or decrease cellular proliferation. This can be explained by the fact that protein of the invention or part thereof is probably involved in rRNA maturation, thus the use of products that inhibit rRNA maturation prevents the formation of functional ribosomes, which leads to an inhibition of protein synthesis. Cells that are unable to synthesis proteins stop to grow and ultimately die due to the fact that they are unable to regenerate proteins. One preferred embodiment of the invention pertains to the use of the protein of the invention or part thereof to develop these antagonists, which are added to samples or materials as a "cocktail" in association with other antimicrobial substances to stop and/or prevent proliferation of undesired contaminants. For example the protein of the invention or part thereof may be used to inhibit the proliferation of undesired bacteria and or viruses in *in vitro* cultures. In another preferred embodiment of the invention the protein or part thereof could be used to develop antagonists that could be administered to patients suffering from viral and or bacterial infection

particularly viral infections by viruses such as HIV and HCV. This could for example be accomplished by targeting the antagonists to cells infected by the virus or directly to bacteria. Once inside these cells the antagonist will inhibit or at least decrease protein synthesis resulting in an inhibition or a decrease in bacterial and/or viral replication. In yet another preferred embodiment of the invention the protein or part thereof could be used to develop antagonists that could be administered to patients in order to inhibit abnormal and/or unregulated cellular proliferation found in diseases such as cancers, psoriasis, Systemic lupus erythematosus (SLE), arthritis, endometriosis, enteropathy in immunodeficiency virus infection, venous eczema (inducing connective tissue sclerosis in lipodermatosclerosis and causing the reduced reepithelialization tendency in venous ulcers), chronic irritant contact dermatitis (CICD), adult polycystic kidney disease (APKD), ichthyosis, cholesteatoma.

Protein of SEQ ID NOs:303 (internal designation number 187-31-0-0-F12-CS) and 275 (internal designation number 145-59-2-0-A7-CS)

The 148-amino-acid long protein of SEQ ID NO:303, encoded by the cDNA of SEQ ID NO:62, found in fetal kidney and highly expressed in this organ, is homologous to the human RNA-associated protein HSCP250 (SPTREMBLNEW SPTREMBL SWISSPROT accession number AAF36170 and GENESEQP accession number Y84433). In addition, this protein displays significant homology to the ribosomal L27 protein of *D. melanogaster* (GENPEPT GENPEPTNEW accession number AE003576) and to the 50S ribosomal L27 protein of *E.coli* (SWISSPROT accession number P02427). The protein of SEQ ID NO:303 has a putative signal peptide, from amino acid position 13 to 27. According to the PFAM program, the protein of the invention also presents a ribosomal L27 protein signature in position 31 to 81. Amino acid residues in position 64 to 78 are highly similar to the consensus pattern: G-X-[LIVM](2)-X-R-Q-R-G-X(5)-G, where X is any amino acid (the motif found in the protein of SEQ ID NO:303 is G-X-I-I-X-T-Q-R-H-X(5)-G). Potential phosphorylation sites exist in positions 32, 38, 47 (S amino residues), 60 (Y amino residue), 69 and 141 (T amino residues). One of them, the T residue in position 69, is embedded in the ribosomal L27 protein signature described above.

The protein of SEQ ID NO:275, encoded by the cDNA of SEQ ID NO:34, is a 94-amino-acid long variant of the SEQ ID NO:303 protein. While the first 81 amino acid residues of protein of SEQ ID NO:275 are strictly homologous to the first 81 amino acid residues of protein of SEQ ID NO:303, the 13 subsequent amino-acids are different. In addition to the putative signal peptide (position 13 to 27), the ribosomal L27 protein signature (position 64 to 78), and phosphorylation sites (positions 32, 38, 47, 60 and 69), the protein of SEQ ID NO:275 also displays a candidate membrane-spanning segment in position 74 to 94.

Ribosomal protein L27 is one of the proteins of the large ribosomal subunit. L27 belongs to a family of ribosomal proteins which, on the basis of sequence similarities, includes: eubacterial

L27, plant chloroplast L27 (nuclear-encoded), algal chloroplast L27 and yeast mitochondrial YmL2 (gene MRPL2 or MRP7). Among the different ribosomal L27 proteins characterized so far, the one of *E.coli* is probably the best studied. Protein L27 is one of the smallest and the most basic polypeptides in *E.coli* ribosome. Techniques like the measurement of protein exposure by hot tritium bombardment have shown that L27 of the large subunit is well exposed on the surface of the *E.coli* 70S ribosome (Agafonov *et al.*, Proc. Natl. Acad. Sci. 94:12892-12897 (1997)). Chemical and UV-crosslinking studies have demonstrated that L27 is closely associated with domain V of the 23S rRNA, a region that comprises part of the peptidyl transferase center (Osswald *et al.*, Nucleic Acids Res. 18:6755-6760 (1990)). Direct evidence for the presence of L27 at the peptidyl transferase center was obtained through the use of derivatives of tRNA^{Phe} containing photoreactive azidonucleotides within the 3'-terminal ACCA_{OH} sequence (Wower *et al.*, Proc. Natl. Acad. Sci. 86:5232-5236 (1989)). Analysis of a mutant *E.coli* strain in which the *rpmA* gene, which encodes L27, was replaced by a Kanamycin marker, has suggested that L27 contributes to peptide bond formation by facilitating the proper placement of the acceptor end of the A-site tRNA at the peptidyl transferase center (Wower *et al.*, J. Biol. Chem. 273:19847-19852 (1998)). Further, recent studies conducted by Thiede and collaborators have precisely determined RNA-protein contact sites in the 50S ribosomal subunit of *E.coli* (Thiede *et al.*, Biochem. J. 334:39-42 (1998)), showing that Lys-71 and Lys-74 of L27 interact with U-2334 of the 23S rRNA.

It is believed that the proteins of SEQ ID NOs:303 and 275 are human RNA-associated proteins. Preferred polypeptides of SEQ ID NO:303 are polypeptides comprising the amino acids from positions 13 to 27, 64 to 78 and amino acid residues in positions 32, 38, 47, 60, 69 and 141. It is believed that the protein of SEQ ID NO:275 is a 94 amino acid long variant of the 148 amino acid residues protein of SEQ ID NO:303. Preferred polypeptides of SEQ ID NO:275 are polypeptides comprising the amino acids from positions 13 to 27, 64 to 78, 74 to 94 and amino acid residues in positions 32, 38, 47, 60, and 69. Other preferred polypeptides of the invention are fragments of SEQ ID NO:303 or 275 having any of the biological activities described herein.

One embodiment of the present invention involves the use of the present proteins and nucleic acids to specifically identify cells from the kidney, especially from the fetal kidney. Such cells can be detected by virtue of their strong expression of the protein of the invention, and can thus be detected using any standard method for detecting protein expression or activity, including methods involving antibodies, specific nucleic acids, or any other detectable molecule that specifically binds to the polypeptides or polynucleotides of the invention. An ability to specifically detect kidney cells is useful, e.g. for determining the identity of tumor cells as well as for the identification of specific cell types and tissues for, e.g. histological analyses.

In another embodiment of the present invention, the present proteins are used as a component of in vitro eukaryotic translation systems. Such systems represent a widely used tool for protein production with many academic and industrial applications. Similarly, inhibitors of the

protein of the invention, e.g. antibodies or dominant negative forms of the protein, can be used to inhibit in vitro translation systems, e.g. to specifically stop a translation reaction involving a eukaryotic cell extract.

In another embodiment, the proteins of SEQ ID NO:303 or 275 can be used to bind to
5 nucleic acids, preferably RNA, alone or in combination with other substances. For example, the proteins of the invention or part thereof can be added to a sample containing RNAs in optimum conditions for binding, and allowed to bind to RNAs. In a preferred such embodiment, the proteins of the invention or part thereof may be used to purify mRNAs, for example to specifically isolate RNA, e.g. from a specific cell type or from cells grown under particular conditions. Such RNAs
10 could then be reverse transcribed and cloned, could be analyzed for relative expression analyses, etc. In addition, such methods may be used to specifically remove RNA from a sample, for example during the purification of DNA. To carry out any of these methods, the proteins of the invention or part thereof may be bound to a chromatographic support, either alone or in combination with other RNA binding proteins, to form an affinity chromatography column. A
15 sample containing a mixture of nucleic acids to purify is then run through the column. Immobilizing the proteins of the invention or part thereof on a support is particularly advantageous for embodiments in which the method is to be practiced on a commercial scale. This immobilization facilitates the removal of RNAs from the batch of resin-coupled protein after binding, and allows subsequent re-use of the protein. Immobilization of the proteins of the
20 invention or part thereof can be accomplished, for example, by inserting any matrix binding domain in the protein according to methods known to those skilled in the art. The resulting fusion product including the proteins of the invention or part thereof is then covalently, or by any other means, bound to a protein, carbohydrate or matrix (such as gold, "Sephadex" particles, polymeric surfaces).

Still another embodiment of the invention relates to methods of preparing antibodies
25 directed against the proteins of the invention or part thereof. Such antibodies may be used, e.g., in co-immunoprecipitation experiments to separate and purify RNAs associated with the proteins of the invention. To accomplish this, in a sample containing a mixture of nucleic acids, antibodies directed against the protein of the invention may be added in association with protein A or protein G sepharose beads. Immunoprecipitation conditions are well known to those skilled in the art.

30 The invention further relates to methods and compositions used to modify the proteins of the invention. In a preferred embodiment, K amino-acids of the proteins of the invention are substituted for other basic amino-residues (R residues), as some of these K residues seem to be crucial for RNA interactions (Thiede et al., Biochem. J. 334:39-42 (1998)). Conversely, R residues of the proteins of the invention may be substituted for K residues. These substitutions are predicted
35 to change the specificity and/or the affinity of the proteins of the invention for RNA molecules. Another preferred embodiment may be to perform post-translational modifications of the proteins of the invention, notably at the level of the putative phosphorylation sites described above in positions

32, 38, 47, 60, 69 and 141 of SEQ ID NO:303. By adding negative charges to the proteins of the invention, these phosphorylation sites may modulate the affinity of the protein for RNA molecules. Phosphorylation of T residue in position 69 is of great interest, as it is embedded in the ribosomal L27 protein signature.

5 In another preferred embodiment, the proteins of the invention or part thereof may be used to visualize RNAs, when the polypeptides are linked to an appropriate fusion partner, or is detected by probing with an antibody.

Another embodiment of the present invention relates to methods and compositions using the proteins of the invention, or part thereof, to associate specific mRNAs to the inner face of lipidic bilayers of liposomes in order to further introduce these mRNAs into the cytoplasm of eukaryotic cells. For example, as described above, the protein of the invention of SEQ ID NO:275 displays both a candidate membrane-spanning segment in position 74 to 94 (at its very carboxy-terminal part), and a ribosomal L27 protein signature in position 64 to 78. Moreover SEQ ID NO:275 is an RNA binding protein. Preferably, specific mRNAs are first associated with the protein of the invention and the RNA/protein complex formed in that way is then mixed with liposomes according to methods known to those skilled in the art. These liposomes are added to an *in vitro* culture of eukaryotic cells. In vivo, such a method might treat and/or prevent disorders linked to dysregulation of gene transcription such as cancer and other disorders relating to abnormal cellular differentiation, proliferation, or degeneration.

20 In another embodiment, the present proteins and nucleic acids can be used to modulate the rate of cell growth in vitro or in vivo. Studies in *Drosophila* have shown that a decrease in ribosome function results in a significant inhibition of cell growth. Accordingly, compounds that inhibits the expression or function of the proteins of the invention can be used to inhibit the growth rate of cells, and can thus be used, e.g. in the treatment or prevention of diseases or conditions associated with excessive cell growth, such as cancer or inflammatory conditions. Such compounds include, but are not limited to, antibodies, antisense molecules, dominant negative forms of the proteins, and any heterologous compounds that inhibit the expression or the activity of the proteins.

Protein of SEQ ID NO:269 (internal designation 116-115-2-0-F8-CS)

30 The protein of SEQ ID NO:269, encoded by the cDNA SEQ ID NO:28, shows homology with the mink whale ribonuclease A (Emmens M., et al., Biochem. J. 157:317-323(1976)) a member of the pancreatic ribonuclease family. In addition, the protein of the invention exhibits 2 membrane spanning segments, the first from amino acid positions 1-21, the second from amino acid positions 179-199. The cDNA SEQ ID NO:28 is composed of 3 exons. Exon 1 is encoded by nucleotides 1-225, exon 2 by nucleotides 226-288, and exon 3 by nucleotides 289-597. The protein of the invention is highly expressed in the testis.

Ribonucleases are proteins that catalyze the hydrolysis of phosphodiester bonds in RNA chains. Pancreatic ribonucleases are pyrimidine-specific ribonucleases present in high quantity in the pancreas of a number of mammalian taxa and of a few reptiles. In addition to their function in the hydrolysis of RNA, ribonucleases have evolved to support a variety of other physiological activities. Such activities include anti-parasite, anti-bacterial, anti-virus, and anti-neoplastic activities, as well as, in some cases, promoting neurotoxicity and angiogenesis. For example, bovine seminal ribonuclease is anti-neoplastic (Laceetti, P. et al. (1992) *Cancer Res.* 52: 4582-4586), and some frog ribonucleases display both anti-viral and anti-neoplastic activity (Youle, R. J. et al. (1994) *Proc. Natl. Acad. Sci. USA* 91: 6012-6016; Mikulski, S. M. et al. (1990) *J. Natl. Cancer Inst.* 82: 151-152; and Wu, Y. -N. et al. (1993) *J. Biol. Chem.* 268: 10686-10693). In addition, angiogenin is a tRNA-specific ribonuclease which binds actin on the surface of endothelial cells for endocytosis and is then translocated to the nucleus where it promotes endothelial invasiveness required for blood vessel formation (Moroianu, J. and Riordan, J. F. (1994) *Proc. Natl. Acad. Sci. USA* 91: 1217-1221). Further, eosinophil-derived neurotoxin (EDN) and eosinophil cationic protein (ECP) are related ribonucleases which possess neurotoxicity (Beintema, J. J. et al. (1988) *Biochemistry* 27: 4530-4538; Ackerman, S. J. (1993) In Makino, S. and Fukuda, T., *Eosinophils: Biological and Clinical Aspects*. CRC Press, Boca Raton, Fla., pp 33-74). ECP also exhibits cytotoxic, anti-parasitic, and anti-bacterial activities. Finally, an EDN-related ribonuclease, RNase k6, is expressed in normal human monocytes and neutrophils, suggesting a role for this ribonuclease in host defense (Rosenberg, H. F. and Dyer, K. D. (1996) *Nuc. Acid. Res.* 24: 3507-3513).

It is believed that the protein of SEQ ID NO:269 is a ribonuclease, and is thus capable of hydrolyzing ribonucleic acids, and is involved in the a number of processes including defense against infection and neoplasia, as well as in neurotoxicity and angiogenesis. Preferred polypeptides of the invention are any fragments of SEQ ID NO:269 having any of the biological activities described herein. The ribonuclease activity of the protein of the invention or part thereof may be assayed using any assay known to those skilled in the art, including those described in US patent 5,866,119.

In one embodiment, the present polynucleotides and polypeptides are used to specifically detect testis tissue and cells derived from the testis, as the present protein is overexpressed in this tissue. For example, the protein of the invention or part thereof may be used to synthesize specific antibodies using any technique known to those skilled in the art. Such tissue-specific antibodies may then be used to identify tissues of unknown origin, such as in forensic samples, differentiated tumor tissue that has metastasized to foreign bodily sites, etc., or to differentiate different tissue types in a tissue cross-section using immunochemistry.

The present invention relates to methods and compositions using the protein of the invention or part thereof to hydrolyze one or several substrates, preferably nucleic acids, more

preferably RNA, alone or in combination with other substances. For example, the protein of the invention or part thereof is added to a sample containing a substrate(s) in conditions amenable to enzyme activity, and the protein thus catalyzes the hydrolysis of the substrate(s).

In a preferred embodiment, the protein of the invention or part thereof may be used to
5 remove contaminating RNA in a biological sample, alone or in combination with other nucleases. In a more preferred embodiment, the protein of the invention or part thereof is used to remove contaminating RNA from DNA preparations, to remove RNA templates prior to second strand synthesis and prior to analysis of in vitro translation products. In one such embodiment, the protein of the invention or part thereof is added to a biological sample as a "cocktail" along with other
10 nucleases. The advantage of using a cocktail of hydrolytic enzymes is that one is able to hydrolyze a wide range of substrates without knowing the specificity of any of the enzymes, or even the identity of all of the substrates. Such cocktails of nucleases are commonly used in molecular biology assays, for example to remove unbound RNA in RNase protection assays. Using a cocktail of hydrolytic enzymes also protects a sample from a wide range of future unknown RNA
15 contaminants from a vast number of sources. For example, the protein of the invention or part thereof is added to samples where contaminating substrates are undesirable. Alternatively, the protein of the invention or part thereof may be bound to a chromatographic support, either alone or in combination with other hydrolytic enzymes, using techniques well known in the art, to form an affinity chromatography column. A sample containing the undesirable substrate is run through the
20 column to remove the substrate. Immobilizing the protein of the invention or part thereof on a support is particularly advantageous for those embodiments in which the method is to be practiced on a commercial scale. This immobilization facilitates the removal of the enzyme from the batch of product and subsequent reuse of the enzyme. Immobilization of the protein of the invention or part thereof can be accomplished, for example, by inserting a cellulose-binding domain in the protein.
25 One of skill in the art will understand that other methods of immobilization could also be used and are described in the available literature. Alternatively, the same methods may be used to identify new substrates.

In another embodiment, the protein of the invention or part thereof may be used to decontaminate or disinfect samples infected by undesirable parasite, bacteria and/or viruses using
30 any of the methods known to those skilled in the art including those described in Youle et al, (1994), supra; Mikulski et al (1990) supra, Wu et al (1993). In a preferred embodiment, the protein is used to eliminate RNA viruses from a sample or in a patient.

In another embodiment, the present invention relates to compositions and methods using the protein of the invention or part thereof to selectively kill cells. The protein of the invention or part
35 thereof is linked to a recognition moiety capable of binding to a chosen cell, such as lectins, receptors or antibodies, thereby generating cell-specific cytotoxic reagents as described in US Patent No. 5,955,073, the disclosure of which is herein incorporated in its entirety.

In another embodiment, the protein of the invention or part thereof is used in the diagnosis, prevention and/or treatment of neoplastic disorders. In one such embodiment, cancer can be treated or prevented in a patient by increasing the activity of the present protein in the patient, particularly within neoplastic or hyperplastic cells within the patient. For example, a polynucleotide encoding the protein of the invention or part thereof, a polynucleotide encoding the protein, or a compound that causes an increase in the expression or activity of the protein, can be administered to the patient, or to cells derived from the patient, in vivo or ex vivo. Preferably, the protein, polynucleotide, or compound is specifically targeted to the neoplastic or hyperplastic cells, for example by intratumoral injection of the molecule or by linking the molecule to a targeting moiety, such as a tumor cell-specific antibody.

In another embodiment, cancer can be treated or prevented in a patient by inhibiting the expression or activity of the protein of the invention in endothelial cells of the patient, in particular within endothelial cells involved in angiogenesis. Such expression or activity can be inhibited in any of a number of ways, for example using antibodies, antisense sequences, ribozymes, dominant negative forms of the protein, as well as small molecule inhibitors of protein activity or expression.

In another embodiment, the present polynucleotide and polypeptide sequences are used to diagnose cancer in a patient. In a typical such embodiment, a biological sample is obtained from a patient, and the level of the present polypeptides or polynucleotides is detected and compared with a control level, where a difference between the level observed in the patient and the control level indicates the presence of cancer in the patient.

In another embodiment, the present protein is inhibited within cells of a mammal in order to protect cells of the mammal against RNase-associated neurotoxicity. In a typical such embodiment, the level of the protein is detected within the cells of the patient, where an elevated level of the protein, particularly within neurons, indicates a risk for neurotoxicity. The level of the expression or activity of the protein is subsequently inhibited using any standard method, such as antibodies, antisense molecules, ribozymes, dominant negative forms of the protein, or any other compounds that inhibit the expression or activity of the protein. Preferably, such inhibitors are specifically directed to the neurons of the mammal.

Protein of SEQ ID NO: 390 (internal designation 116-118-4-0-A8-CS)

The present inventors have provided a new gene and protein described in SEQ ID No 149 and 390 respectively, belonging to the carbonic anhydrase (CA) family, more particularly the alpha-CA family. This novel alpha-CA related gene is located on the human chromosome 17q24 region.

The Carbonic anhydrases (EC 4.2.1.1) (CA) are zinc metalloenzymes which catalyze the reversible hydration of carbon dioxide. Nine different active Alpha-carbonic anhydrases (alpha-CA) that catalyze the hydration reaction have been found, as well as at least two alpha-CA-related enzymes. All known carbonic anhydrases from the animal kingdom are alpha-CAs, as opposed to

beta- and gamma-CAs, which are also zinc containing enzymes but are unrelated by sequence. The protein of SEQ ID No. 390 displays significant homology to the pfam Carbonic anhydrase domains amino acids between 20- to 59 of the protein, in particular the motif Gly-Ser-Glu-His in position 45 to 48 of the protein which has been found to be highly conserved in a multi-alignment published by Lovejoy et al. 1998 (Genomics 54, 484-493). The chromosomal localization was found by BLAST alignment with a sequence mapping to chromosome 17 (genbank genomic fragment with accession number AC002090) and another alignment with a genomic fragment (accession number AF064854) which maps the gene in the 17q24 region. The polypeptide of SEQ ID No. 390 displays particularly high homology to a human protein called carbonic anhydrase-related protein 10 (genbank accession number AB036836, published directly in database by Adachi, K. and Nishimori, I).

The human alpha-CAs contain a highly conserved catalytic site which comprises a zinc coordination polyhedron defining an active site located in a large cone-shaped cavity that extends almost to the center of the alpha-CA molecule. One site of the cavity is formed by hydrophobic residues, which the other side contains hydrophilic residues, including Thr199 and Glu106 (referring to CA II enzyme). The zinc ion is located at the bottom of this cleft, and tetrahedrally coordinated to the imidazoles of three histidine residues (His94, His96, His 119, referring to CA II enzyme) and to a water molecule called the 'zinc water' that ionizes to a hydroxide ion with a pK of about 7. (Sly et al., Ann. Rev. Biochem. 64:375-401 (1995). Studies have shown that the Zn-OH-/Thr199/Glu106 network is important in binding bicarbonate, sulfonamide inhibitors and many anionic inhibitors (Liljas et al., Eur. J. Biochem. 219:1-10).

Improved alpha-CA inhibitors

Of the human alpha-CAs, it has been found that the various isozymes have differing tissue distributions and intracellular localizations. Alpha-CA II for example, is expressed in the cytosol of some cell types in virtually every tissue or organ, while alpha-CA I is expressed in colon and erythrocytes, and alpha-CA IV is expressed on the apical surfaces of epithelial cells of some segments of the nephron, the apical plasma membrane in the lower gastrointestinal tract, and the plasma face of endothelial cells of certain capillary beds. The protein of SEQ ID NO: 390 encoded by the cDNA of SEQ ID NO: 149 has been found by the present inventors to be expressed in testes.

The human alpha-CAs have been found to be involved in a range of important biological functions involving pH regulation, CO₂ and HCO₃⁻ transport, ion transport and water and electrolyte balance. Functions in which alpha-CAs are involved include H⁺ secretion, HCO₃⁻ reabsorption, HCO₃⁻ secretion, bone resorption, and production of aqueous humor, cerebrospinal fluid, gastric acid and pancreatic juice. Of particular medical interest, CA II has been found to be implicated in osteoporosis, as CA II defects have been found to be a cause of inherited osteoporosis, found along with renal tubular acidosis and brain calcification.

CA activity can be determined by well known means. Assays used to characterize CA isozyme activity are provided for example in Khalifah, J. Biol. Chem. 246:2561-73 (1971); Chen et

al, Biochem 32: 7861-65 (1993); Tu et al., J. Biol. Chem. 258:8867-8871 (1986); and Jewel et al., Biochem. 30:1484-1490 (1991).

Many different inhibitors of CA have been identified, and certain CA inhibitors have been developed as medicaments. CA inhibitors are currently a primary treatment for glaucoma, where inhibition of CA activity reduces intra-ocular pressure by inhibiting formation of aqueous humor. Approved CA inhibitors for glaucoma include Acetazolamide (Diamox®), Methazolamide (Neptazane®), Dorzolamide (Trusopt®) and Brinzolamide (Azopt®).

Improved broad-acting CA inhibitors

In certain treatment settings, there is a need for CA inhibitors capable of inhibiting the broad class of CA isozymes so as to inhibit CO₂ hydration activity. Local (topical) use of CA inhibitors has been found advantageous over systemic application for glaucoma, allowing systemic side effects of CA inhibitors to be avoided. However, current CA inhibitors may have limited efficacy in terms of ability to completely inhibit total CA activity. A topical treatment, Dorzolamide (Trosopt®), for example, is an inhibitor of CA II, but only weakly inhibits CA VI. (Hoyng et al., Drugs, 50(3): 411-434 (2000). Inhibitors of CA II may be unable to inhibit CA I or other CAs, which is thought to result in decreased drug efficacy because other CAs can compensate for loss of CA II activity (Sly and Hu, Ann. Rev. Biochem. 64:375-401 (1995)). In one example, CA I is five to six times as abundant as CA II in human erythrocytes, but has only about 15% of the activity. Thus, CA I contributes about 50% of the total CA activity (Dodgon et al., J. Appl. Physiol. 64:1492-80 (1988). Moreover, CA I may have different inhibitor sensitivity profile from CA II, as CA I is less sensitive to sulfonamide inhibitors, for example. CA II and CA IV on the other hand, show significant resistance to inhibition with halide ions in comparison to CA I. (Sly et al, (1995), supra) Thus, a significant amount of residual CA activity in a cell or tissue of interest may be due to other CAs, including the polypeptide of SEQ ID No. 390.

Thus, in one aspect, alpha-CA related nucleic acid and polypeptide may be useful for the identification of compounds capable of inhibiting the alpha-CA-catalyzed reversible hydration reaction. In one aspect, the method is carried out to identify or select CA inhibitors capable of inhibiting the activity of the polypeptide of SEQ ID No. 390. In other aspects, the method is carried out to identify or select CA inhibitors capable of broadly inhibiting the activity of a large number of CA enzymes. The nucleic acid and polypeptide sequences of the invention can be used in computer based drug design or for carrying out binding predictions with candidate CA inhibitors in view of the extensive structural information publicly available for CA enzymes. In preferred embodiments, the nucleic acid and polypeptide of the invention is used in drug screening assays. Assays may be cell based or non-cell based assays. In one embodiment, a nucleotide or polypeptide sequence of the invention is brought into contact with a candidate CA inhibitor (such as a CA II inhibitor), and binding of the candidate inhibitor to the polypeptide of the invention, or the activity of the polypeptide of the invention is detected. Activity of the polypeptide of the invention may be CA

activity, or any other suitable activity possessed by the polypeptide of the invention which may be inhibited by binding of the candidate substance. Assays for detecting hydration of carbon dioxide are well known, and referenced above. In preferred embodiments, a panel of CA isozymes including the polypeptide of the invention are screened against the candidate substance, including one or more
5 enzymes selected from the group consisting of CA I, CA III, CA IV, CA VI, and a CA-RP including but not limited to CA-RP VII, CA-RP X and CA-RP XI. In preferred embodiments, a candidate CA inhibitor is selected according to its ability to broadly inhibit CA isozymes capable of catalyzing the hydration of carbon dioxide. Means to conduct such drug screening assays are well known in the art. In one embodiment drug binding is tested, using means described further herein
10 as well as for example in International Patent Publication No. WO 00/58510, the disclosure of which is incorporated herein by reference in its entirety, particularly the section titled "Methods for screening substances interacting with... polypeptides". Drug binding assays on a large panel of isozymes may also be carried out in high throughput format using commercially available binding assay systems (Graffinity Pharmaceutical Design GmbH, Heidelberg, Germany).

15 The method according to the invention may generally be used to identify or select candidate compounds for the treatment of a disorder characterized by a disorder in pH regulation, CO_2 and HCO_3^- transport, ion transport or water and electrolyte balance.

Improved selective CA inhibitors

In other therapeutic strategies, CA inhibitors are delivered orally. However, systemic
20 delivery may affect CA enzymes present in other tissues or organs leading to harmful side effects. It can be expected that CA II inhibitors may also partially or fully inhibit other CA isozymes, such as CA I, or a CA-related protein (CA-RP) such as CA-RP VIII, X, XI or CA RPTP-beta (Tashian et al., In "*Carbonic Anhydrase: New Horizons*", W.R. Chegwidden, N.D. Carter and Y.H. Edwards, Eds., Birkhauser, Basel); Adachi, K et al., Genbank accession number AB036836; Lovejoy et al.
25 (1998); Peles et al. (1995)) or the CA polypeptide of SEQ ID No 390. CA-RPTP-beta, for example, has a CA domain having no or reduced CA activity but is thought to be involved in the ligand binding or protein complex participation in view of its binding of contactin by an extracellular region. (Peles et al., Cell 82: 251-260 (1995); Tashian et al., (1998), supra). The inhibition of an isozyme such as CA-RPTP-beta or the isozyme of the present invention by systemic treatment using
30 a non-selective drug may result in harmful side effects.

In one example, while oral CA inhibitors for the treatment of glaucoma (eg. acetazolamide) have been effective and without ocular adverse effects, they have shown important systemic effects, including parasthesia of the acra, fatigue, depression, renal stones and gastrointestinal complaints such as nausea and diarrhoea. (Hoyng et al., 2000, supra) Because CA inhibitors are typically used
35 permanently (eg for glaucoma), or over long periods of time, avoiding side effects is particularly important. Selective CA inhibitors capable of inhibiting a CA isozyme of interest to a greater extent than another CA isozyme may thus offer improved means for the treatment of disease.

In one embodiment, the nucleotide and polypeptide sequences of the present invention may be used to design selective CA inhibitors. Studies have also shown that the different alpha-CA have different inhibitor binding properties (Sly et al., (1995), supra), suggesting that it may be possible to provide compounds that inhibit a CA isozyme of interest, such as CA II, while not binding to or inhibiting related enzymes such as the polypeptide of SEQ ID No. 390. The nucleic acid and polypeptide sequences of the invention can be used in computer based drug design or for carrying out binding predictions with candidate CA inhibitors in view of the extensive structural information publicly available for CA enzymes. In preferred embodiments, the nucleic acid and polypeptide of the invention is used in drug screening assays, including both cell based and non cell based assays.

In one embodiment, a nucleotide or polypeptide sequence of the invention is brought into contact with a candidate CA inhibitor (such as a CA II inhibitor), and binding of the candidate inhibitor to the polypeptide of the invention, or the activity of the polypeptide of the invention is detected. Activity of the polypeptide of the invention may be CA activity, or any other suitable activity possessed by the polypeptide of the invention which may be inhibited by binding of the candidate substance. In preferred embodiments, a panel of CA isozymes including the polypeptide of the invention are screened against the candidate substance, including the polypeptide of SEQ ID No 39 and one or more enzymes selected from the group consisting of CA I, CA III, CA IV, CA VI, a CARP including but not limited to CARP VII, CARP X, CARP XI. In preferred embodiments, a candidate CA inhibitor is screened against one or more non-catalytic CA related proteins to eliminate undesired inhibition of these enzymes which may be involved in other important physiological functions. Means to conduct such drug screening assays are well known in the art.

Increasing alpha-CA activity for the treatment of alpha-CA deficiency disease

The polypeptide of the invention may also be used as a source of CA activity, such as for the treatment of disease. The defects in carbonic anhydrases are the cause of several diseases, including osteopetrosis (abnormally dense bone) renal tubular acidosis, cerebral calcification and mental retardation. Also, a carbonic anhydrase-related protein is described as being linked to cone-rod retinal dystrophy (Bellinghan et al., 1998, Biochem. Biophys. Res. Comm.: 253, 364-367).

In one aspect, the invention thus involves increasing CA activity by providing increased activity of the polypeptide of SEQ ID No. 390. Increased activity of the polypeptide of SEQ ID No 390 can be provided by any suitable means, as further described herein. Activity may be provided for example by introducing to a host cell or patient a vector containing a nucleotide sequence of SEQ ID No 149, treating said cell with a compound capable of increasing the expression of the polypeptide of the invention and/or treating a cell or patient directly with a polypeptide of SEQ ID No 390. In preferred embodiments, the polypeptide of the invention comprises at least one amino acid substitution, deletion or insertion. In one aspect, such amino acid changes are preferably in the catalytic site; preferably said amino acid changes involve the substitution, deletion or insertion of a

His residue and preferably said amino acid changes increase the CO₂ hydration activity of the polypeptide of the invention.

Metal ion biosensors

In further aspects, metal ion biosensors can be designed based on the polypeptide of SEQ

- 5 ID No 390. Determination of metal ion concentrations in complex media such as serum, cell cytoplasm as well as for example seawater are important analytical functions that require high degrees of sensitivity and selectivity.

Biosensors may be particularly useful in detecting metal ion fluxes in and between cells.

- Such biosensors may exploit metal-binding ability of the polypeptide of the invention, as described
10 by Thompson et al., who have developed such biosensors based on the CA enzyme (CA II). Such biosensors are useful in the detection of metal ion flux for example in the central nervous system. Zinc-containing neurons found throughout the mammalian cerebral cortex, striatum and amygdalar nuclei have been shown to release their zinc in a depolarization- and calcium-dependent fashion in vitro and in vivo. This zinc release has been suggested to act as a trans-synaptic neuromodulator :
15 which has in turn been linked to excitotoxic neuronal cell death. CA based biosensors developed by Thompson et al. showed that zinc is present and can be detected in extracellular medium from neurons. (Thompson et al, J. Neurosci Methods 96:35-45 (2000)).

- Biosensors based on CA have been shown to be extremely selective, detecting Cu at subpicomolar levels, which is of sensitivity that might be achieved with mass spectrometric
20 techniques. Sensors based on the CA II isozyme have been shown to detect Zn and Cu at picomolar levels, and Cd, Co and Ni at nanomolar levels. (Thompson et al., Anal. Biochem. 267:185-195 (1999)). CA based biosensors have also demonstrated selectivity over potential interferents in biological systems at mM levels in extracellular fluids, such as Mg and Ca. (Thompson et al. (2000), supra).

- 25 Biosensors based on the polypeptide of the invention are based on the high selectivity and sensitivity of CA isozymes for zinc. Because the binding of Zn in the active site of the enzyme affects the enzyme's ability to bind a CA inhibitor, it is possible to use a CA inhibitor that exhibits a detectable change upon binding to the polypeptide of the invention to detect the fraction of polypeptide bound to the inhibitor, and therefore bound to Zn. The fraction of polypeptide with
30 bound Zn in turn is determined by the concentrations of free Zn and the polypeptide of the invention, and the dissociation constant for zinc.

- In one example, binding of the CA inhibitor to the polypeptide of the invention is detected by using a fluorescent inhibitor, whereby the inhibitor shows a detectable change in fluorescence emission wavelength of polarization upon binding to the polypeptide of the invention. In one
35 example, a fluorescent sulfonamide is used, such as the fluorophore ABD-N (Thompson et al. (2000), supra).

Engineered CA enzymes

CA isozymes have been shown to have differing levels of catalytic activity and efficiency. In preferred embodiments, particularly for treatments which involve providing the increased activity of the polypeptide of SEQ ID No 390 or for use in metal ion biosensors, the polypeptide of the invention may be modified for increased CO₂ hydration and/or zinc binding.

- 5 In particular, studies have been carried out characterizing residues important for maximal CA activity, allowing CA isozymes to be designed having desired levels of activity. Important structural elements in CA isozymes for zinc binding, CO₂ hydration activity and stability are reviewed in Lindskog, Pharmacol. Ther. 74(1):1-20 (1997) and Sly (1995), supra. In one example, studies of CA III showed that changing the Phe198 residue to a Leu198 residue (as in CAII)
- 10 resulted in a 25 fold increase in activity. (Chen et al., (1993), supra). Catalysis has also been greatly increased in CA II by replacing the Thr200 residue with His, as is normally found in CA I enzymes. Most dramatically, a CA-related protein (CA-RP) which in its native form was missing important residues at the catalytic site and had no detectable CO₂ hydration activity at all was rendered an active CA by only two point mutations. (Sjoblom et al., FEBS Lett. 398: 322-325(1996)).
- 15 Thus, in embodiments where the polypeptide of the invention is used to provide a source of CO₂ hydration or for its zinc binding properties, it is advantageous to modify the polypeptide of the invention by introducing at least one amino acid substitution, deletion or insertion. In one aspect, such amino acid changes are preferably in the catalytic site; preferably said amino acid changes involve the substitution, deletion or insertion of a His residue and preferably said amino acid
- 20 changes increase the CO₂ hydration activity of the polypeptide of the invention. Optimal amino acid changes can be determined by the skilled artisan, particularly in view of sequence comparisons which can be carried out with the many well-characterized CA isozymes.

Protein of SEQ ID NO:252 (internal designation 105-089-3-0-G10-CS)

- The protein of SEQ ID NO:252 is encoded by the cDNA of SEQ ID NO:11. Accordingly,
- 25 it will be appreciated that all characteristics and uses of the polypeptide of SEQ ID NO:252 described throughout the present application also pertain to the polypeptide encoded by the human cDNA of clone 105-089-3-0-G10-CS. In addition, it will be appreciated that all characteristics and uses of the nucleic acid of SEQ ID NO:11 described throughout the present application also pertain to the human cDNA of clone 105-089-3-0-G10-CS. It is over represented in fetal brain.

- 30 The protein of SEQ ID NO:252 encoded by the cDNA of SEQ ID NO:11 is distributed primarily in the prostate and salivary gland. The protein of SEQ ID NO:252 is homologous to sequences described in PCT publication WO9827205-A2 (which describes a protein that was isolated from a human adult salivary gland cDNA library), PCT publication WO9839446-A2, PCT publication WO9839446-A2. The disclosures of each of the preceding PCT publications is
- 35 incorporated herein by reference in their entireties.

The protein of SEQ ID NO:252 is also homologous to a polypeptide described in PCT publication WO9835229-A1, the disclosure of which is incorporated herein by reference in its entirety. WO9835229-A1 describes a peptide of 27 amino acid residues that corresponds to 23/27 of a portion of the protein of SEQ ID NO:252 (amino acid 20-46). This corresponds to 85% identity with conserved changes (3 out of 4) yielding a 96% homology.

The protein described in WO 9835229 was identified in reflex tears that were collected from 12 non-contact lens wearing male and female humans. Reflex tears were stimulated by gently rubbing the nasal mucosa with a cotton wool tipped bud. Two different batches were collected from two different groups and examined by analytical and preparative 2-dimensional electrophoresis. After separation in the second dimension and transfer to PVDF membranes, identified protein spots (by 0.1% (w/v) Coomassie Blue) were loaded into a membrane-compatible Hewlett-Packard cartridge. Sequencing was conducted with a Model G1005A (Hewlett-Packard, CA) sequenator. One of the proteins identified migrated at 25 kDa and was revealed to have 5 isoforms of different pI. Two of these were N-terminally sequenced and gave the sequence of the above peptide with a pI of 5.0 and 4.4. The different isoforms indicate that this protein undergoes post-translational modifications, including sialylation or acylation. The presence of these isoforms in different degrees could reflect the disease status of the individual. Accordingly, one embodiment of the present invention relates to the detection or diagnosis of disease by determining the activity or level of the protein of SEQ ID NO:252 or a polynucleotide encoding the protein of SEQ ID NO:252 in an individual. For example, detection of the secreted protein of SEQ ID NO:252 in an individual may be accomplished non-invasively by measuring protein levels in bodily fluids into which the protein is secreted, such as tears and saliva. Such methods may be employed both in humans and in animals. It is probable that after the signal peptide is cleaved, the protein of SEQ ID NO:252 is secreted into bodily fluids including tears and probably saliva.

The protein of SEQ ID NO:252 can also be used for the screening of non-ocular diseases, by analyzing tears for marker proteins, particularly indicative of cancer and genetic disease. In addition, an altered chromatographic profile (e.g. 2D gel) of the isoforms of the protein of SEQ ID NO:252 may also indicate the disease state of an individual. For example, the levels of marker proteins in relation to the protein of SEQ ID NO:252 may be determined to evaluate whether the individual is suffering from a disease. Alternatively, tears may be analyzed for the levels of different isoforms of the protein of SEQ ID NO:252 to determine whether the pattern of such isoforms is indicative of disease.

The protein of SEQ ID NO:252 or fragments thereof may also be used as a lubricant or cleansing agent for the eyes. This protein can be included in contact lenses washing and storage solutions. This protein can also be useful as an ingredient in eye washing solutions (e.g. eye drops) used for everyday redness or healing after surgical/laser intervention. For example, the protein may be used to reduce eye inflammation. Alternatively, anti-bacterial properties may be exploited by

including the protein of SEQ ID NO:252 or fragments thereof in solutions, creams or ointments for the eyes, as well as creams or ointments in general for external applications.

Accordingly, the present invention includes the use of the protein of SEQ ID NO:252, fragments comprising at least 5, 8, 10, 12, 15, 20, 25, 30, 35, 40, 50, 60, 75, 100, 150, or 200 consecutive amino acids thereof, or fragments having a desired biological activity to treat or ameliorate a condition in an individual. In such embodiments, the protein of SEQ ID NO:252, or a fragment thereof, is administered to an individual in whom it is desired to increase or decrease any of the activities of the protein of SEQ ID NO:252. The protein of SEQ ID NO:252 or fragment thereof may be administered directly to the individual or, alternatively, a nucleic acid encoding the protein of SEQ ID NO:252 or a fragment thereof may be administered to the individual. Alternatively, an agent which increases the activity of the protein of SEQ ID NO:252 may be administered to the individual. Such agents may be identified by contacting the protein of SEQ ID NO:252 or a cell or preparation containing the protein of SEQ ID NO:252 with a test agent and assaying whether the test agent increases the activity of the protein. For example, the test agent may be a chemical compound or a polypeptide or peptide.

Alternatively, the activity of the protein of SEQ ID NO:252 may be decreased by administering an agent which interferes with such activity to an individual. Agents which interfere with the activity of the protein of SEQ ID NO:252 may be identified by contacting the protein of SEQ ID NO:252 or a cell or preparation containing the protein of SEQ ID NO:252 with a test agent and assaying whether the test agent decreases the activity of the protein. For example, the agent may be a chemical compound, a polypeptide or peptide, an antibody, or a nucleic acid such as an antisense nucleic acid or a triple helix-forming nucleic acid.

In one embodiment, the invention relates to methods and compositions using the protein of the invention or part thereof as a marker protein to selectively identify the source of a sample as, for example, saliva or tears, or to distinguish between two or more possible sources of a sample on the basis of the level of the protein of SEQ ID NO:252 in the sample. For example, the protein of SEQ ID NO:252 or fragments thereof may be used to generate antibodies using any techniques known to those skilled in the art, including those described therein. Such antibodies may then be used to identify tissues of unknown origin, for example, forensic samples, differentiated tumor tissue that has metastasized to foreign bodily sites, or to differentiate different tissue types in a tissue cross-section using immunochemistry. In such methods a sample is contacted with the antibody, which may be detectably labeled, under conditions which facilitate antibody binding. The level of antibody binding to the test sample is measured and compared to the level of binding to control cells from saliva or tears or tissues other than saliva or tears to determine whether the test sample is from saliva or tears. Alternatively, the level of the protein of SEQ ID NO:252 in a test sample may be measured by determining the level of RNA encoding the protein of SEQ ID NO:252 in the test sample. RNA levels may be measured using nucleic acid arrays or using techniques such as in situ

hybridization, Northern blots, dot blots or other techniques familiar to those skilled in the art. If desired, an amplification reaction, such as a PCR reaction, may be performed on the nucleic acid sample prior to analysis. The level of RNA in the test sample is compared to RNA levels in control cells from saliva or tears or tissues other than saliva or tears to determine whether the test sample is
5 from saliva or tears.

In another embodiment, antibodies to the protein of the invention or part thereof may be used for detection, enrichment, or purification of cells expressing the protein of SEQ ID NO:252, including using methods known to those skilled in the art. For example, an antibody against the protein of SEQ ID NO:252 or a fragment thereof may be fixed to a solid support, such as a
10 chromatography matrix. A preparation containing cells expressing the protein of SEQ ID NO:252 is placed in contact with the antibody under conditions which facilitate binding to the antibody. The support is washed and then the cells are released from the support by contacting the support with agents which cause the cells to dissociate from the antibody.

In another embodiment of the present invention, the protein of SEQ ID NO:252 or a
15 fragment thereof thereof may be used to diagnose disorders associated with altered expression of the protein of SEQ ID NO:252. In such techniques, the level of the protein of SEQ ID NO:252 in an ill individual is measured using techniques such as those described herein. The level of the protein of 252 in the ill individual is compared to the level in normal individuals to determine whether the individual has a level of the protein of SEQ ID NO:252 which is indicative of disease.

20 Protein of SEQ ID NO:308 (internal designation 187-41-0-0-i21-CS)

The protein of SEQ ID NO:308 is encoded by the cDNA of SEQ ID NO:67. Accordingly, it will be appreciated that all characteristics and uses of the polypeptide of SEQ ID NO:308 described throughout the present application also pertain to the polypeptide encoded by the human cDNA of clone 187-41-0-0-i21-CS. In addition, it will be appreciated that all characteristics and
25 uses of the nucleic acid of SEQ ID NO:67 described throughout the present application also pertain to the human cDNA of clone 187-41-0-0-i21-CS.

The protein of SEQ ID NO:308 is highly homologous to human secreted protein nf87_1 from PCT publication WO 9935252-A2 (the disclosure of which is incorporated herein by reference in its entirety), to amino acids 26-129 of the human secreted protein SEQ ID NO:441 from PCT
30 publication WO 9906548-A2 (the disclosure of which is incorporated herein by reference in its entirety), and to amino acids 26-114 of human secreted protein SEQ ID NO:439 from PCT publication WO 9906548-A2, the disclosure of which is incorporated herein by reference in its entirety. Thus, the protein of the invention appears to be a polymorphic variant of nf87_1. Since most of the proteins with high homology to the sequence of the invention have longer 5' termini, it
35 is conceivable that the protein of the invention is a truncated/spliced variant of these proteins.

The protein of SEQ ID NO:308 was identified among the cDNAs from a library constructed from brain. Tissue distribution analysis through a BLAST analysis of databases shows that mRNA encoding this protein was found primarily in kidney, liver, and cancerous prostate.

The protein of SEQ ID NO:308 has chemical and structural homology to human interferon-inducible (IFI) protein isoforms p27 (63%), HIFI (50% identity), and to interferon-induced protein 6-16 precursor (IFI-6-16, 36%). Furthermore, the protein of the invention has structural homology (40% identity) to the human erythropoietin (EPO) primary response gene, EPRG3pt from PCT publication WO 9906063-A2, the disclosure of which is incorporated herein by reference in its entirety. Thus, the present invention relates to nucleic acid and amino acid sequences of a novel IFI protein and to the use of these sequences in the diagnosis, study, prevention and treatment of disease.

The protein of SEQ ID NO:308 comprises 105 amino acids. From the amino acid alignments and the hydrophobicity plots, it has a predicted signal peptide sequence spanning residues 31-43 and two predicted transmembrane domains spanning residues 17-37, and 48-68. Accordingly, one embodiment of the present invention is a polypeptide comprising the signal peptide and/or one or more of the transmembrane domains.

Interferons (IFNs) are a part of the group of intercellular messenger proteins known as cytokines. α -IFN is the product of a multigene family of at least 16 members, whereas b-IFN is the product of a single gene. α - and β -IFNs are also known as type I IFNs. Type I IFNs are produced in a variety of cells types. Biosynthesis of type I IFNs is stimulated by viruses and other pathogens, and by various cytokines and growth factors. γ -IFN, also known as type II IFN, is produced in T-cells and natural killer cells. Antigens to which the organism has been sensitized stimulate biosynthesis of type II IFN. Both α - and γ -IFNs are immunomodulators and anti-inflammatory agents, activating macrophages, T-cells and natural killer cells.

IFNs are part of the body's natural defense to viruses and tumors. They exert these defenses by affecting the function of the immune system and by direct action on pathogens and tumor cells. IFNs mediate these multiple effects in part by inducing the synthesis of many cellular proteins. Some interferon-inducible (IFI) genes are induced equally well by α -, β - and γ -IFNs. Other IFI genes are preferentially induced by the type I or by the type II IFNs. The various proteins produced by IFI genes possess antitumor, antiviral and immunomodulatory functions. The expression of tumor antigens in cancer cells is increased by α -IFN, and renders the cancer cells more susceptible to immune rejection. The IFI proteins synthesized in response to viral infections are known to inhibit viral functions such as cell penetration, uncoating, RNA and protein synthesis, assembly and release (Hardman JG et al 25 (1996) The Pharmacological Basis of Therapeutics, McGraw-Hill, New York NY pp 1211-1214, the disclosure of which is incorporated herein by reference in its entirety). Type II IFN stimulates expression of major histocompatibility complex (MHC) proteins and is thus used in immune response enhancement.

The IFI gene known as 6-16 encodes an mRNA, which is highly induced by type I IFNs in a variety of human cells (Kelly JM et al (1986) EMBO J 5:1601-1606, the disclosure of which is incorporated herein by reference in its entirety). After induction, 6-16 mRNA constitutes as much as 0.1% of the total cellular mRNA. The 6-16 mRNA is present at only very low levels in the
5 absence of type I IFN, and is only weakly induced by type II IFN. The 6-16 mRNA encodes a hydrophobic protein of 130 amino acids. The first 20 to 23 amino acids comprise a putative signal peptide. Protein 6-16 has at least two predicted transmembrane regions culminating in a negatively charged C-terminus.

The p27 gene encodes a protein with 41% amino acid sequence identity to the 6-16 protein.
10 The p27 gene is expressed in some breast tumor cell lines and in a gastric cancer cell line. In other breast tumor cell lines, in the HeLa cervical cancer cell line, and in fetal lung fibroblasts, p27 expression occurs only upon α -IFN induction. In one breast tumor cell line, p27 is independently induced by estradiol and by IFN (Rasmussen UB et al (1993) Cancer Res 53:4096-4101, the disclosure of which is incorporated herein by reference in its entirety). Expression of p27 was
15 analyzed in 21 primary invasive breast carcinomas, 1 breast cancer bone metastasis, and 3 breast fibroadenomas. High levels of p27 were found in about one-half of the primary carcinomas and in the bone metastasis, but not in the fibroadenomas. These observations suggest that certain breast tumors may produce high levels of, or have increased sensitivity to, type I IFN as compared to other breast tumors (Rasmussen UB et al, supra). In addition, the p27 gene expressed at significant levels
20 in normal tissues including colon, stomach and lung, but not expressed in placenta, kidney, liver or skin. (Rasmussen UB et al, supra).

The small hydrophobic IFI gene products may contribute to viral resistance. A hepatitis-C virus (HCV)-induced gene, 130-51, was isolated from a cDNA library prepared from chimpanzee liver during the acute phase of the infection. The protein product of this gene has 97% identity to
25 the human 6-16 protein (Kato T et al (1992) Virology 190:856-860, the disclosure of which is incorporated herein by reference in its entirety). The authors of the preceding paper suggest that HCV infection actively induces IFN expression, which in turn induces expression of IFI genes including 130-51. The IFI proteins synthesized in response to viral infections are known to inhibit viral functions such as penetration, uncoating, RNA or protein synthesis, assembly or release. The
30 130-51 protein may inhibit one or more of these functions in HCV. A particular virus may be inhibited in multiple functions by IFI proteins. In addition, the principle inhibitory effect exerted by IFI proteins differs among the virus families (Hardman JG, supra, p 1211, the disclosure of which is incorporated herein by reference).

The HIFI protein (PCT publication WO 9812223-A2, the disclosure of which is
35 incorporated herein by reference in its entirety) is a human sequence identified among cDNAs from a library constructed from human neonatal kidney. Northern blot analysis using LIFESEQTM database (Incyte Pharmaceuticas, Palo Alto, CA) shows that HIFI mRNA was found only in

neonatal kidney. The HIFI protein consists of 104 amino acids and has 55%, 45%, and 46% amino acid sequence identity to p27, 6-16 and 130-51, respectively.

Based on the chemical and structural homology between the protein of SEQ ID NO:308 and the small hydrophobic IFI proteins from human and chimpanzee, it is believed that the protein of
5 SEQ ID NO:308 is synthesized when interferons are produced in infections, inflammation, autoimmune diseases etc. Interferons are produced in response to various cytokines and growth factors, in viral infections, inflammation, autoimmune diseases, and cancers. Accordingly, the protein of SEQ ID NO:308 or fragments thereof may be used in diagnosis and treatment of diseases such as, but not limited to, autoimmune disorders such as rheumatoid arthritis, Graves disease,
10 systemic lupus erythematosus, autoimmune hepatitis, Wegener's granulomatosis, sarcoidosis, polyarthritis, pemphigus, pemphigoid, erythema multiform, Sjogren's syndrome, inflammatory bowel disease, multiple sclerosis, myasthenia gravis keratitis, scleritis, Type I diabetes, insulin-dependent diabetes mellitus, Lupus Nephritis, and allergic encephalomyelitis; proliferative disorders including various forms of cancer such as leukemias, lymphomas (Hodgkins and non-Hodgkins),
15 sarcomas, melanomas, adenomas, carcinomas of solid tissue, hypoxic tumors, squamous cell carcinomas of the mouth, throat, larynx, and lung, genitourinary cancers such as cervical and bladder cancer, hematopoietic cancers, head and neck cancers, and nervous system cancers, benign lesions such as papillomas, atherosclerosis, angiogenesis; viral infections, in particular HCV and HIV infections, as well as other pathogen-induced infections (e.g. leishmania).

20 The protein of SEQ ID NO:308 or fragments thereof may also be used to treat conditions associated with inflammation or immune impairment (e. g. reumathoid and osteo arthritis and AIDS).

Another embodiment of the present invention relates to the use of the protein of SEQ ID NO:308 or fragments thereof to treat and/or prevent the ill-effect of bacterial infection during
25 pregnancy in mammals, such as spontaneous abortion and maternal death. In a preferred embodiment, the protein of the invention may be used to counteract the effects of the bacterial endotoxin lipopolysaccharide (LPS). The methods for using such compositions is described in Dziegielewska and Andersen, Biol. Neonate, 74:372-5 (1998), the disclosure of which is incorporated herein by reference in its entirety.

30 Furthermore, the protein of SEQ ID NO:308 or fragments thereof are useful as a reagent for analyzing the control of gene expression by interferons and other cytokines in both normal and diseased cells. The protein of the SEQ ID NO:308 or fragments thereof may be used to identify specific molecules with which it binds such as agonists, antagonists or inhibitors.

Another embodiment of the present invention relates to methods of using the protein of
35 SEQ ID NO:308 or fragments thereof to identify and/or quantify cytokines of the interferon family as well as other cytokines such as IL10 and tumor antigens, which may interact with the protein of the invention.

The protein of SEQ ID NO:308 or fragments thereof may also be included in pharmaceutical preparations for treating cancer or prevention/treatment of other diseases associated with changes in expression of the protein of the invention. In another embodiment of the present invention, the protein of SEQ ID NO 308 or fragments thereof is used to inhibit and/or modulate the effect of cytokines and related molecule such as Il-2, TNF alpha, CTLA4, CD28, and others, by preventing the binding of the endogenous cytokine to their natural receptors, thereby blocking cell proliferation or inhibitory signals generated by the ligand-receptor binding event.

The protein of SEQ ID NO:308 or fragments thereof is useful to correct defects in in vivo models of disease such as autoimmune, inflammation and tumor models, by injecting the protein either intra peritoneally intravenously, subcutaneously or directly in the diseased tissue.

The DNA encoding the protein of SEQ ID NO:308 or fragments thereof is useful in diagnostic assays for conditions/diseases associated with expression of the protein of the invention. The diagnostic assay is useful to distinguish between absence, presence, and excess expression of the protein of the invention and to monitor regulation of levels of the protein of the invention during therapeutic intervention. The DNA may also be incorporated into effective eukaryotic expression vectors and directly targeted to a specific tissue, organ, or cell population for use in gene therapy to treat the above mentioned conditions, including tumors and/or to correct disease- or genetic-induced defects in any of the above mentioned proteins including the protein of the invention. The DNA may also be used to design antisense sequences and ribozymes, which can be administered to modify gene expression in tumor and pathogen-infected cells and to influence expression of cytokines and growth factors. In vivo delivery of genetic constructs into subjects can be developed to the point of targeting specific cell types, such as tumor where expression of the protein of the invention may be affected or is modulating the expression and/or activity of other proteins such as cytokines, growth factors, their receptors and/or tumor antigens. It is also useful to detect unknown upstream sequences (e. g. promoters and regulatory elements) by standard techniques and for research into the control of gene expression by interferons and other cytokines, as well as growth and transcription factors in normal and diseased cells. Hybridization probes are useful to detect DNA encoding the protein of the invention (or closely related molecules) in biological samples, and for mapping the naturally occurring genomic sequence to a particular chromosome/chromosome region. The DNA may be used to generate and/or treat in vivo animal models of disease, including susceptibility or resistance to infection, inflammation, tumors and autoimmune conditions, as well as tumor therapy, based on vaccine, knock-out and transgene technologies.

Antibodies against the protein of SEQ ID NO:308 or fragments thereof are useful for the diagnosis of conditions and diseases associated with its expression and to quantify the protein of the invention (e. g. in assays to monitor patients during therapeutic intervention). Antibodies specific for the protein may include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, Fab fragments produced by a Fab expression library. Neutralizing antibodies are especially

preferred for diagnostics and therapeutics. Diagnostic assays for the protein of the invention include methods utilizing the antibody and a label to detect the protein of the invention in human body fluids or extracts of cells or tissues.

The protein of the invention and its catalytic or immunogenic fragments or oligopeptides thereof, can be used for screening therapeutic compounds in any variety of drug screening techniques including high throughput. Methods which may be used to quantitate the expression of the nucleotide or protein of the invention include, but are not limited to, polymerase chain reaction (PCR), RT-PCR, RNase protection, Northern and western blotting, enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), fluorescent activated cell sorting (FACS), immunoprecipitation, and chromatography.

Under conditions of significant blood loss, EPO therapy, or both, iron-restricted erythropoiesis is evident. However, intravenous or oral iron therapy has substantial drawbacks. Moreover, traditional biochemical markers of storage iron in patients with anemia of chronic disease are unhelpful in the assessment of iron status (Lawrence T et al (2000) Blood 96:823-833, the disclosure of which is incorporated herein by reference in its entirety). As the protein of SEQ ID NO:308 bears homology to the human erythropoietin (EPO) primary response gene, EPRG3pt, it may be used to promote red blood cell formation or to monitor the value of safer intravenous iron preparations in patients with blood loss anemia, particularly those undergoing EPO therapy.

The hydrophobic IFI protein of SEQ ID NO:308 or fragments thereof may be used to diagnose conditions associated with its induction. For example, the protein of SEQ ID NO:308 or fragments thereof may be useful in the diagnosis and treatment of tumors, viral infections, inflammation, or conditions associated with impaired immunity, anemia of chronic blood loss or chronic disease, hemochromatosis, and EPO therapy. Furthermore, this protein may be used for investigating the control of gene expression by IFNs and other cytokines, as well as hormones and growth factors, in normal and diseased cells.

The protein of SEQ ID NO:308 or fragments thereof is useful to correct defects in in vivo models of disease such as autoimmune, inflammation, anemia, iron-overload and tumor models, by injecting the protein either intra peritoneally intravenously, subcutaneously or directly in the diseased tissue.

In addition, the protein of SEQ ID NO:308 is structurally related to other proteins having homology and/or structural similarity with human p27 (Rasmussen, U.B., et al., 1993, Cancer Research 53:4096-4101, the disclosure of which is incorporated herein by reference). Accordingly, the protein of brain, fetal brain, kidney, fetal kidney, or colon may be used to regulate the proliferation of EPO-dependent cells or the growth and development of erythroid and other hematopoietic lineages.

The protein of SEQ ID NO:308 or fragments thereof, or polynucleotides encoding the protein of SEQ ID NO:308 or fragments thereof, may be used to treat or ameliorate anemia of

chronic disease and chronic renal failure, polycythemia, cancer, AIDS, drug- and phlebotomy-induced anemias, hemochromatosis, erythropoiesis mediated by EPO therapy, and other conditions associated with altered activity or levels of the protein of SEQ ID NO:308.

In another embodiment, the present invention relates to methods for identifying agonists and antagonists/inhibitors using the protein of SEQ ID NO:308 or fragments thereof, and treating conditions with the identified compounds. In a still further aspect, the invention relates to diagnostic assays for detecting diseases associated with inappropriate levels or activity of the protein of SEQ ID NO:308. In still another embodiment of the invention relates to the use of the protein SEQ ID NO:308, fragments thereof or the DNA encoding the protein of SEQ ID NO:308 or fragments thereof to monitor the value of iron therapy in patients undergoing EPO therapy, or experiencing blood loss, or both.

The DNA encoding the protein of SEQ ID NO:308 or fragments thereof is useful in diagnostic assays for conditions/diseases associated with abnormal expression of the protein of SEQ ID NO:308. The diagnostic assay is useful to distinguish between absence, presence, and excess expression of the protein of the invention and to monitor regulation of levels of the protein of the invention during therapeutic intervention. The DNA may also be incorporated into effective eukaryotic expression vectors and directly targeted to a specific tissue, organ, or cell population for use in gene therapy to treat the above mentioned conditions, including tumors and/or to correct disease- or genetic-induced defects in any of the above mentioned proteins including the protein of the invention. The DNA may also be used to design antisense sequences and ribozymes, which can be administered to modify gene expression in tumor and pathogen-infected cells and to influence expression of cytokines, hormones and growth factors. In vivo delivery of genetic constructs into subjects can be developed to the point of targeting specific cell types, such as tumor where expression of the protein of the invention may be affected or is modulating the expression and/or activity of other proteins such as cytokines, growth factors, their receptors and/or tumor antigens. It is also useful to detect unknown upstream sequences (e. g. promoters and regulatory elements) by standard techniques and for research into the control of gene expression by interferons and other cytokines, as well as growth and transcription factors in normal and diseased cells. Hybridization probes are useful to detect DNA encoding the protein of the invention (or closely related molecules) in biological samples, and for mapping the naturally occurring genomic sequence to a particular chromosome/chromosome region. The DNA may be used to generate and/or treat in vivo animal models of disease, including susceptibility or resistance to infection, tumors, autoimmune conditions, anemia and iron-overload, as well as tumor therapy, based on vaccine, knock-out and transgene technologies.

Antibodies against the protein of SEQ ID NO:308 are useful for the diagnosis of conditions and disease associated with its expression and to quantify the protein of the invention (e. g. in assays to monitor patients during therapeutic intervention). Antibodies specific for the protein may

include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, Fab fragments produced by a Fab expression library. Neutralizing antibodies are especially preferred for diagnostics and therapeutics. Diagnostic assays for the protein of SEQ ID NO:308 include methods utilizing the antibody and a label to detect the protein of the invention in human body fluids or

5 extracts of cells or tissues.

The protein of SEQ ID NO:308 and its catalytic or immunogenic fragments or oligopeptides thereof, can be used for screening therapeutic compounds in any variety of drug screening techniques including high throughput. Methods which may be used to quantitate the expression of the nucleotide or protein of the invention include, but are not limited to, polymerase chain reaction
10 (PCR), RT-PCR, RNase protection, Northern blotting, enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), fluorescent activated cell sorting (FACS), immunoprecipitation, and chromatography.

Accordingly, the present invention includes the use of the protein of SEQ ID NO:308, fragments comprising at least 5, 8, 10, 12, 15, 20, 25, 30, 35, 40, 50, 60, 75, 100, 150, or 200
15 consecutive amino acids thereof, or fragments having a desired biological activity to treat or ameliorate a condition in an individual. For example, the condition may be cancer, including breast cancer, viral infection, bacterial infection, inflammation, autoimmune disorders, rheumatoid arthritis, Graves disease, systemic lupus erythematosus, autoimmune hepatitis, Wegener's granulomatosis, sarcoidosis, polyarthritis, pemphigus, pemphigoid, erythema multiform, Sjogren's
20 syndrome, inflammatory bowel disease, multiple sclerosis, myasthenia gravis keratitis, scleritis, Type I diabetes, insulin-dependent diabetes mellitus, Lupus Nephritis, and allergic encephalomyelitis; proliferative disorders including various forms of cancer such as leukemias, lymphomas (Hodgkins and non-Hodgkins), sarcomas, melanomas, adenomas, carcinomas of solid tissue, hypoxic tumors, squamous cell carcinomas of the mouth, throat, larynx, and lung,
25 genitourinary cancers such as cervical and bladder cancer, hematopoietic cancers, head and neck cancers, and nervous system cancers, benign lesions such as papillomas, atherosclerosis, angiogenesis; viral infections, in particular HCV and HIV infections, as well as other pathogen-induced infections (e. g. leishmania).

In such embodiments, the protein of SEQ ID NO:308, or a fragment thereof, is
30 administered to an individual in whom it is desired to increase or decrease any of the activities of the protein of SEQ ID NO:308. The protein of SEQ ID NO:308 or fragment thereof may be administered directly to the individual or, alternatively, a nucleic acid encoding the protein of SEQ ID NO:308 or a fragment thereof may be administered to the individual. Alternatively, an agent which increases the activity of the protein of SEQ ID NO:308 may be administered to the
35 individual. Such agents may be identified by contacting the protein of SEQ ID NO:308 or a cell or preparation containing the protein of SEQ ID NO:308 with a test agent and assaying whether the

test agent increases the activity of the protein. For example, the test agent may be a chemical compound or a polypeptide or peptide.

Alternatively, the activity of the protein of SEQ ID NO:308 may be decreased by administering an agent which interferes with such activity to an individual. Agents which interfere with the activity of the protein of SEQ ID NO:308 may be identified by contacting the protein of SEQ ID NO:308 or a cell or preparation containing the protein of SEQ ID NO:308 with a test agent and assaying whether the test agent decreases the activity of the protein. For example, the agent may be a chemical compound, a polypeptide or peptide, an antibody, or a nucleic acid such as an antisense nucleic acid or a triple helix-forming nucleic acid.

10 In one embodiment, the invention relates to methods and compositions using the protein of the invention or part thereof as a marker protein to selectively identify the source of a sample as, for example, brain, kidney, liver, or cancerous prostate, or to distinguish between two or more possible sources of a sample on the basis of the level of the protein of SEQ ID NO:308 in the sample. For example, the protein of SEQ ID NO:308 or fragments thereof may be used to generate antibodies using any techniques known to those skilled in the art, including those described therein. Such antibodies may then be used to identify tissues of unknown origin, for example, forensic samples, differentiated tumor tissue that has metastasized to foreign bodily sites, or to differentiate different tissue types in a tissue cross-section using immunochemistry. In such methods a sample is contacted with the antibody, which may be detectably labeled, under conditions which facilitate antibody binding. The level of antibody binding to the test sample is measured and compared to the level of binding to control cells from brain, kidney, liver, or cancerous prostate or tissues other than brain, kidney, liver, or cancerous prostate to determine whether the test sample is from brain, kidney, liver, or cancerous prostate. Alternatively, the level of the protein of SEQ ID NO:308 in a test sample may be measured by determining the level of RNA encoding the protein of SEQ ID NO:308 in the test sample. RNA levels may be measured using nucleic acid arrays or using techniques such as in situ hybridization, Northern blots, dot blots or other techniques familiar to those skilled in the art. If desired, an amplification reaction, such as a PCR reaction, may be performed on the nucleic acid sample prior to analysis. The level of RNA in the test sample is compared to RNA levels in control cells from brain, kidney, liver, or cancerous prostate or tissues other than brain, kidney, liver, or cancerous prostate to determine whether the test sample is from brain, kidney, liver, or cancerous prostate.

In another embodiment, antibodies to the protein of the invention or part thereof may be used for detection, enrichment, or purification of cells expressing the protein of SEQ ID NO:308, including using methods known to those skilled in the art. For example, an antibody against the protein of SEQ ID NO:308 or a fragment thereof may be fixed to a solid support, such as a chromatography matrix. A preparation containing cells expressing the protein of SEQ ID NO:308 is placed in contact with the antibody under conditions which facilitate binding to the antibody. The

support is washed and then the cells are released from the support by contacting the support with agents which cause the cells to dissociate from the antibody.

In another embodiment of the present invention, the protein of SEQ ID NO:308 or a fragment thereof thereof may be used to diagnose disorders associated with altered expression of the protein of SEQ ID NO:308. In such techniques, the level of the protein of SEQ ID NO:308 in an ill individual is measured using techniques such as those described herein. The level of the protein of SEQ ID NO:308 in the ill individual is compared to the level in normal individuals to determine whether the individual has a level of the protein of SEQ ID NO:308 which is associated with disease.

10 Protein of SEQ ID NOs:289 and 307 (internal designations 175-1-3-0-E5-CS.cor and 187-39-0-0-k12-CS)

The protein of SEQ ID NO:289 is encoded by the cDNA of SEQ ID NO:48. Accordingly, it will be appreciated that all characteristics and uses of the polypeptide of SEQ ID NO:289 described throughout the present application also pertain to the polypeptide encoded by the human cDNA of clone 175-1-3-0-E5-CS. In addition, it will be appreciated that all characteristics and uses of the nucleic acid of SEQ ID NO:48 described throughout the present application also pertain to the human cDNA of clone 175-1-3-0-E5-CS.

The protein of the invention consists of 130 amino acids. From the amino acid alignments and the hydrophobicity plots, it has a predicted signal peptide sequence spanning residues 8-20 and four predicted transmembrane domains spanning residues 2-24, 42-61, 70-90 and 99-119. Accordingly, some embodiments of the present invention relate to polypeptides comprising the signal peptide and/or one or more of the transmembrane domains.

The protein of SEQ ID NO:289 encoded by the cDNA of SEQ ID NO:48 is homologous to SEQ ID NO: 4199 from EP 1033401-A2 (the disclosure of which is incorporated herein by reference in its entirety), a human secreted protein. Another protein, SEQ ID NO:307, encoded by the cDNA of SEQ ID NO:66, is a polymorphic variant of the protein of SEQ ID NO:289, and shares all of the herein-described functions and uses.

The present invention relates to a novel protein identified among the cDNAs from a library constructed from salivary gland, and to the use of the nucleic acid and amino acid sequences disclosed herein in the study, diagnosis, prevention, and treatment of disease. Tissue distribution analysis predicted by BLAST on databases shows that mRNA encoding this protein was found primarily in brain and fetal brain, with lower amounts in kidney, fetal kidney and colon.

Interferons (IFNs) are a part of the group of intercellular messenger proteins known as cytokines. α -IFN is the product of a multigene family of at least 16 members, whereas b-IFN is the product of a single gene. α - and β -IFNs are also known as type I IFNs. Type I IFNs are produced in a variety of cells types. Biosynthesis of type I IFNs is stimulated by viruses and other pathogens,

and by various cytokines and growth factors. γ -IFN, also known as type II IFN, is produced in T-cells and natural killer cells. Antigens to which the organism has been sensitized stimulate biosynthesis of type II IFN. Both α - and γ -IFNs are immunomodulators and anti-inflammatory agents, activating macrophages, T-cells and natural killer cells.

5 IFNs are part of the body's natural defense to viruses and tumors. They exert these defenses by affecting the function of the immune system and by direct action on pathogens and tumor cells. IFNs mediate these multiple effects in part by inducing the synthesis of many cellular proteins. Some interferon-inducible (IFI) genes are induced equally well by α -, β - and γ -IFNs. Other IFI genes are preferentially induced by the type I or by the type II IFNs. The various proteins produced
10 by IFI genes possess antitumor, antiviral and immunomodulatory functions. The expression of tumor antigens in cancer cells is increased by α -IFN, and renders the cancer cells more susceptible to immune rejection. The IFI proteins synthesized in response to viral infections are known to inhibit viral functions such as cell penetration, uncoating, RNA and protein synthesis, assembly and release (Hardman JG et al 25 (1996) The Pharmacological Basis of Therapeutics, McGraw-Hill,
15 New York NY pp 1211-1214, the disclosure of which is incorporated herein by reference in its entirety). Type II IFN stimulates expression of major histocompatibility complex (MHC) proteins and is thus used in immune response enhancement.

The protein of SEQ ID NO:289 is a small hydrophobic protein having chemical and structural homology to human interferon-inducible (IFI) protein isoforms 6-16 (97% identity), HIF1
20 (44%), and p27 (33%), as well as 130-51, the chimpanzee homolog of 6-16 - (97%). Thus, the protein of SEQ ID NO:289 and the nucleic acid encoding it are polymorphic variants of 6-16 or the gene encoding 6-16. The protein of SEQ ID NO:289, fragments thereof, or nucleic acids encoding the protein of SEQ ID NO:289 or fragments thereof may be used in the diagnosis, study, prevention and treatment of disease as described below.

25 The IFI gene known as 6-16 encodes an mRNA, which is highly induced by type I IFNs in a variety of human cells (Kelly JM et al (1986) EMBO J 5:1601-1606, the disclosure of which is incorporated herein by reference in its entirety). After induction, 6-16 mRNA constitutes as much as 0.1% of the total cellular mRNA. The 6-16 mRNA is present at only very low levels in the absence of type I IFN, and is only weakly induced by type II IFN. The 6-16 mRNA encodes a hydrophobic
30 protein of 130 amino acids. The first 20 to 23 amino acids comprise a putative signal peptide. Protein 6-16 has at least two predicted transmembrane regions culminating in a negatively charged C-terminus.

The p27 gene encodes a protein with 41% amino acid sequence identity to the 6-16 protein. The p27 gene is expressed in some breast tumor cell lines and in a gastric cancer cell line. In other
35 breast tumor cell lines, in the HeLa cervical cancer cell line, and in fetal lung fibroblasts, p27 expression occurs only upon α -IFN induction. In one breast tumor cell line, p27 is independently induced by estradiol and by IFN (Rasmussen UB et al (1993) Cancer Res 53:4096-4101, the

disclosure of which is incorporated herein by reference in its entirety). Expression of p27 was analyzed in 21 primary invasive breast carcinomas, 1 breast cancer bone metastasis, and 3 breast fibroadenomas. High levels of p27 were found in about one-half of the primary carcinomas and in the bone metastasis, but not in the fibroadenomas. These observations suggest that certain breast tumors may produce high levels of, or have increased sensitivity to, type I IFN as compared to other breast tumors (Rasmussen UB et al, supra). In addition, the p27 gene expressed at significant levels in normal tissues including colon, stomach and lung, but not expressed in placenta, kidney, liver or skin. (Rasmussen UB et al, supra).

The small hydrophobic IFI gene products may contribute to viral resistance. A hepatitis-C virus (HCV)-induced gene, 130-51, was isolated from a cDNA library prepared from chimpanzee liver during the acute phase of the infection. The protein product of this gene has 97% identity to the human 6-16 protein (Kato T et al (1992) Virology 190:856-860, the disclosure of which is incorporated herein by reference in its entirety). The authors of this paper suggest that HCV infection actively induces IFN expression, which in turn induces expression of IFI genes including 130-51. The IFI proteins synthesized in response to viral infections are known to inhibit viral functions such as penetration, uncoating, RNA or protein synthesis, assembly or release. The 130-51 protein may inhibit one or more of these functions in HCV. A particular virus may be inhibited in multiple functions by IFI proteins. In addition, the principle inhibitory effect exerted by IFI proteins differs among the virus families (Hardman JG, supra, p 1211, the disclosure of which is incorporated herein by reference).

The HIFI protein (PCT publication WO 9812223-A2, the disclosure of which is incorporated herein by reference in its entirety) is a human sequence identified among cDNAs from a library constructed from human neonatal kidney. Northern blot analysis using LIFESEQ™ database (Incyte Pharmaceuticas, Palo Alto, CA) shows that HIFI mRNA was found only in neonatal kidney. The HIFI protein consists of 104 amino acids and has 55%, 45%, and 46% amino acid sequence identity to p27, 6-16 and 130-51, respectively.

The hydrophobic IFI proteins of the invention may provide the basis for clinical diagnosis of diseases associated with their induction. These proteins may be useful in the diagnosis and treatment of tumors, viral infections, inflammation, or conditions associated with impaired immunity. Furthermore, these proteins may be used for investigations of the control of gene expression by IFNs and other cytokines in normal and diseased cells.

Based on the chemical and structural homology among the protein of SEQ ID NO:289 and the small hydrophobic IFI proteins from human and chimpanzee, it is believed that the protein of SEQ ID NO:289 is synthesized when interferons are produced in infections, inflammation, autoimmune diseases etc. Interferons are produced in response to various cytokines and growth factors, in viral infections, inflammation, autoimmune diseases, and cancers. Accordingly, the protein of SEQ ID NO:289 or fragments thereof may be used in diagnosis and treatment of diseases

such as, but not limited to, autoimmune disorders such as rheumatoid arthritis, Graves disease, systemic lupus erythematosus, autoimmune hepatitis, Wegener's granulomatosis, sarcoidosis, polyarthritis, pemphigus, pemphigoid, erythema multiform, Sjogren's syndrome, inflammatory bowel disease, multiple sclerosis, myasthenia gravis keratitis, scleritis, Type I diabetes, insulin-
5 dependent diabetes mellitus, Lupus Nephritis, and allergic encephalomyelitis; proliferative disorders including various forms of cancer such as leukemias, lymphomas (Hodgkins and non-Hodgkins), sarcomas, melanomas, adenomas, carcinomas of solid tissue, hypoxic tumors, squamous cell carcinomas of the mouth, throat, larynx, and lung, genitourinary cancers such as cervical and bladder cancer, hematopoietic cancers, head and neck cancers, and nervous system cancers, benign
10 lesions such as papillomas, atherosclerosis, angiogenesis; viral infections, in particular HCV and HIV infections, as well as other pathogen-induced infections (e. g. leishmania).

The protein of SEQ ID NO:289 or fragments thereof may also be used to treat conditions associated with inflammation or immune impairment (e. g. reumathoid and osteo arthritis and AIDS).

15 Another embodiment of the present invention relates to the use of the protein of SEQ ID NO:289 or fragments thereof to treat and/or prevent the ill-effect of bacterial infection during pregnancy in mammals, such as spontaneous abortion and maternal death. In a preferred embodiment, the protein of the invention may be used to counteract the effects of the bacterial endotoxin lipopolysaccharide (LPS). The methods for using such compositions is described in
20 Dziegielewska and Andersen, Biol. Neonate, 74:372-5 (1998), the disclosure of which is incorporated herein by reference in its entirety.

Furthermore, the protein of SEQ ID NO:289 or fragments thereof are useful as a reagent for analyzing the control of gene expression by interferons and other cytokines in both normal and diseased cells. The protein of the SEQ ID NO:289 or fragments thereof may be used to identify
25 specific molecules with which it binds such as agonists, antagonists or inhibitors.

Another embodiment of the present invention relates to methods of using the protein of SEQ ID NO:289 or fragments thereof to identify and/or quantify cytokines of the interferon family as well as other cytokines such as IL-10 and tumor antigens, which may interact with the protein of the invention.

30 The protein of SEQ ID NO:289 or fragments thereof may also be included in pharmaceutical preparations for treating cancer or prevention/treatment of other diseases associated with changes in expression of the protein of the invention. In another embodiment of the present invention, the protein of SEQ ID NO:289 or fragments thereof is used to inhibit and/or modulate the effect of cytokines and related molecule such as Il-2, TNF alpha, CTLA4, CD28, and others, by
35 preventing the binding of the endogenous cytokine to their natural receptors, thereby blocking cell proliferation or inhibitory signals generated by the ligand-receptor binding event.

The protein of SEQ ID NO:289 or fragments thereof is useful to correct defects in in vivo models of disease such as autoimmune, inflammation and tumor models, by injecting the protein either intra peritoneally intravenously, subcutaneously or directly into the diseased tissue.

The DNA encoding the protein of SEQ ID NO:289 or fragments thereof is useful in
5 diagnostic assays for conditions/diseases associated with expression of the protein of the invention. The diagnostic assay is useful to distinguish between absence, presence, and excess expression of the protein of the invention and to monitor regulation of levels of the protein of the invention during therapeutic intervention. The DNA may also be incorporated into effective eukaryotic expression vectors and directly targeted to a specific tissue, organ, or cell population for use in gene therapy to
10 treat the above mentioned conditions, including tumors and/or to correct disease- or genetic-induced defects in any of the above mentioned proteins including the protein of the invention. The DNA may also be used to design antisense sequences and ribozymes, which can be administered to modify gene expression in tumor and pathogen-infected cells and to influence expression of cytokines and growth factors. In vivo delivery of genetic constructs into subjects can be developed
15 to the point of targeting specific cell types, such as tumor where expression of the protein of the invention may be affected or is modulating the expression and/or activity of other proteins such as cytokines, growth factors, their receptors and/or tumor antigens. It is also useful to detect unknown upstream sequences (e. g. promoters and regulatory elements) by standard techniques and for research into the control of gene expression by interferons and other cytokines, as well as growth
20 and transcription factors in normal and diseased cells. Hybridization probes are useful to detect DNA encoding the protein of the invention (or closely related molecules) in biological samples, and for mapping the naturally occurring genomic sequence to a particular chromosome/chromosome region. The DNA may be used to generate and/or treat in vivo animal models of disease, including susceptibility or resistance to infection, inflammation, tumors and autoimmune conditions, as well
25 as tumor therapy, based on vaccine, knock-out and transgene technologies.

Antibodies against the protein of SEQ ID NO:289 or fragments thereof are useful for the diagnosis of conditions and diseases associated with its expression and to quantify the protein of the invention (e. g. in assays to monitor patients during therapeutic intervention). Antibodies specific for the protein may include, but are not limited to, polyclonal, monoclonal, chimeric, single chain,
30 Fab fragments produced by a Fab expression library. Neutralizing antibodies are especially preferred for diagnostics and therapeutics. Diagnostic assays for the protein of the invention include methods utilizing the antibody and a label to detect the protein of the invention in human body fluids or extracts of cells or tissues.

The protein of the invention and its catalytic or immunogenic fragments or oligopeptides
35 thereof, can be used for screening therapeutic compounds in any variety of drug screening techniques including high throughput. Methods which may be used to quantitate the expression of the nucleotide or protein of the invention include, but are not limited to, polymerase chain reaction

(PCR), RT-PCR, RNase protection, Northern and western blotting, enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), fluorescent activated cell sorting (FACS), immunoprecipitation, and chromatography.

Accordingly, the present invention includes the use of the protein of SEQ ID NO:289, fragments comprising at least 5, 8, 10, 12, 15, 20, 25, 30, 35, 40, 50, 60, 75, 100, 150, or 200 consecutive amino acids thereof, or fragments having a desired biological activity to treat or ameliorate a condition in an individual. For example, the condition may be cancer, including breast cancer, viral infection, bacterial infection, inflammation, autoimmune disorders, rheumatoid arthritis, Graves disease, systemic lupus erythematosus, autoimmune hepatitis, Wegener's granulomatosis, sarcoidosis, polyarthritis, pemphigus, pemphigoid, erythema multiform, Sjogren's syndrome, inflammatory bowel disease, multiple sclerosis, myasthenia gravis keratitis, scleritis, Type I diabetes, insulin-dependent diabetes mellitus, Lupus Nephritis, and allergic encephalomyelitis; proliferative disorders including various forms of cancer such as leukemias, lymphomas (Hodgkins and non-Hodgkins), sarcomas, melanomas, adenomas, carcinomas of solid tissue, hypoxic tumors, squamous cell carcinomas of the mouth, throat, larynx, and lung, genitourinary cancers such as cervical and bladder cancer, hematopoietic cancers, head and neck cancers, and nervous system cancers, benign lesions such as papillomas, atherosclerosis, angiogenesis; viral infections, in particular HCV and HIV infections, as well as other pathogen-induced infections (e. g. leishmania).

In such embodiments, the protein of SEQ ID NO:289, or a fragment thereof, is administered to an individual in whom it is desired to increase or decrease any of the activities of the protein of SEQ ID NO:289. The protein of SEQ ID NO:289 or fragment thereof may be administered directly to the individual or, alternatively, a nucleic acid encoding the protein of SEQ ID NO:289 or a fragment thereof may be administered to the individual. Alternatively, an agent which increases the activity of the protein of SEQ ID NO:289 may be administered to the individual. Such agents may be identified by contacting the protein of SEQ ID NO:289 or a cell or preparation containing the protein of SEQ ID NO:289 with a test agent and assaying whether the test agent increases the activity of the protein. For example, the test agent may be a chemical compound or a polypeptide or peptide.

Alternatively, the activity of the protein of SEQ ID NO:289 may be decreased by administering an agent which interferes with such activity to an individual. Agents which interfere with the activity of the protein of SEQ ID NO:289 may be identified by contacting the protein of SEQ ID NO:289 or a cell or preparation containing the protein of SEQ ID NO:289 with a test agent and assaying whether the test agent decreases the activity of the protein. For example, the agent may be a chemical compound, a polypeptide or peptide, an antibody, or a nucleic acid such as an antisense nucleic acid or a triple helix-forming nucleic acid.

In one embodiment, the invention relates to methods and compositions using the protein of the invention or part thereof as a marker protein to selectively identify the source of a sample as, for example, brain, fetal brain, kidney, fetal kidney, or colon, or to distinguish between two or more possible sources of a sample on the basis of the level of the protein of SEQ ID NO:289 in the sample. For example, the protein of SEQ ID NO:289 or fragments thereof may be used to generate antibodies using any techniques known to those skilled in the art, including those described therein. Such antibodies may then be used to identify tissues of unknown origin, for example, forensic samples, differentiated tumor tissue that has metastasized to foreign bodily sites, or to differentiate different tissue types in a tissue cross-section using immunochemistry. In such methods a sample is contacted with the antibody, which may be detectably labeled, under conditions which facilitate antibody binding. The level of antibody binding to the test sample is measured and compared to the level of binding to control cells from brain, fetal brain, kidney, fetal kidney, or colon or tissues other than brain, fetal brain, kidney, fetal kidney, or colon to determine whether the test sample is from brain, fetal brain, kidney, fetal kidney, or colon. Alternatively, the level of the protein of SEQ ID NO:289 in a test sample may be measured by determining the level of RNA encoding the protein of SEQ ID NO:289 in the test sample. RNA levels may be measured using nucleic acid arrays or using techniques such as in situ hybridization, Northern blots, dot blots or other techniques familiar to those skilled in the art. If desired, an amplification reaction, such as a PCR reaction, may be performed on the nucleic acid sample prior to analysis. The level of RNA in the test sample is compared to RNA levels in control cells from brain, fetal brain, kidney, fetal kidney, or colon or tissues other than brain, fetal brain, kidney, fetal kidney, or colon to determine whether the test sample is from brain, fetal brain, kidney, fetal kidney, or colon.

In another embodiment, antibodies to the protein of the invention or part thereof may be used for detection, enrichment, or purification of cells expressing the protein of SEQ ID NO:289, including using methods known to those skilled in the art. For example, an antibody against the protein of SEQ ID NO:289 or a fragment thereof may be fixed to a solid support, such as a chromatography matrix. A preparation containing cells expressing the protein of SEQ ID NO:289 is placed in contact with the antibody under conditions which facilitate binding to the antibody. The support is washed and then the cells are released from the support by contacting the support with agents which cause the cells to dissociate from the antibody.

In another embodiment of the present invention, the protein of SEQ ID NO:289 or a fragment thereof thereof may be used to diagnose disorders associated with altered expression of the protein of SEQ ID NO:289. In such techniques, the level of the protein of SEQ ID NO:289 in an ill individual is measured using techniques such as those described herein. The level of the protein of SEQ ID NO:289 in the ill individual is compared to the level in normal individuals to determine whether the individual has a level of the protein of SEQ ID NO:289 which is associated with disease.

Protein of SEQ ID NO:268 (internal designation 116-111-4-0-B3-CS)

The protein of SEQ ID NO:268 is encoded by the cDNA of SEQ ID NO:27. Accordingly, it will be appreciated that all characteristics and uses of the polypeptide of SEQ ID NO:268 described throughout the present application also pertain to the polypeptide encoded by the human
5 cDNA of clone 116-111-4-0-B3-CS. In addition, it will be appreciated that all characteristics and uses of the nucleic acid of SEQ ID NO:27 described throughout the present application also pertain to the human cDNA of clone 116-111-4-0-B3-CS. The protein of the invention is found to be expressed in testis and lungs.

The protein of SEQ ID NO:268 encoded by the extended cDNA SEQ ID NO: 27 is a
10 splicing variant of XAGE-1, a member of the CT antigen family overexpressed in Ewing sarcoma (Liu, X. F., L. J. Helman, et al. (2000). Cancer Res 60(17): 4752-5, the disclosures of which are incorporated by reference herein in their entireties). In addition, the protein of SEQ ID NO:268 also shows strong homology at the COOH end with PAGE4, another member of the CT antigen family (Brinkmann, U., G. Vasmatzis, et al. (1999) Cancer Res 59(7): 1445-8, the disclosure of which is
15 incorporated herein by reference in its entirety).

The cDNA SEQ ID NO:27 is composed of 5 exons. Exon 1 lies between nucleotides 1-245, exon2 lies between nucleotides 246-370, exon 3 lies between nucleotides 371-512, exon 4 lies between nucleotides 513-639, and exon 5 lies between nucleotides 640-762 . Exons 2 to 5 of cDNA SEQ ID NO:27 are shared in part with XAGE-1. However, since the initiation codon of SEQ ID
20 NO: 27 is located in intron1 of XAGE-1, there is a frameshift in the alignment of the 2 molecules. Exon 1 of SEQ ID NO:27 lies between nucleotides 110-234 of XAGE-1, exon 2 of SEQ ID NO:27 lies between nucleotides 235-376 of XAGE-1, exon 3 of SEQ ID NO:27 lies between nucleotides 377-503 of XAGE-1, and exon 4 of SEQ ID NO:27 lies between nucleotides 504-526 of XAGE-1.

XAGE-1 is overexpressed in sarcoma and alveolar rhabdomyosarcoma and is also highly
25 expressed in normal testis (Liu, X. F., L. J. Helman, et al. (2000). Cancer Res 60(17): 4752-5, the disclosure of which is incorporated herein by reference in its entirety). In addition XAGE-1 share homology with PAGE-4 (Brinkmann, U., G. Vasmatzis, et al. (1999) Cancer Res 59(7): 1445-8, the disclosure of which is incorporated herein by reference in its entirety) at the COOH end.

CT antigens are a distinct class of differentiation antigens that are expressed by cancers
30 arising in nonessential normal tissues such as prostate, breast, and ovary (G. Vasmatzis et al., Proc. Natl. Acad. Sci. USA, 95: 300-304, 1998, the disclosure of which is incorporated herein by reference in its entirety) and that have a restricted pattern of expression in normal tissues. This class of antigens are presented on the surface of tumor cells and are recognized by cytolytic T cells, leading to lysis. The extent to which these antigens have been studied, has been via cytolytic T cell
35 characterization studies, in vitro i.e., the study of the identification of the antigen by a particular cytolytic T cell ("CTL" hereafter) subset. The subset proliferates upon recognition of the presented tumor rejection antigen, and the cells presenting the antigen are lysed. Characterization studies have

identified CTL clones which specifically lyse cells expressing the antigens. Examples of this work may be found in Levy et al., Adv. Cancer Res. 24: 1-59 (1977); Boon et al., J. Exp. Med. 152: 1184-1193 (1980); Brunner et al., J. Immunol. 124: 1627-1634 (1980); Maryanski et al., Eur. J. Immunol. 124: 1627-1634 (1980); Maryanski et al., Eur. J. Immunol. 12: 406-412 (1982); Palladino et al., Canc. Res. 47: 5074-5079 (1987), the disclosures of which are incorporated herein by reference in their entireties.

Some thoroughly studied CT antigens are MAGE, BAGE, GAGE and LAGE, others have been added including PAGE, XAGE, most of them located on chromosome X. Brinkmann et al reported the identification of three new members of the GAGE/ PAGE family, termed XAGEs. XAGE-1 and XAGE-2 are expressed in Ewing's sarcoma, rhabdomyosarcoma, a breast cancer, and a germ cell tumor.

It is believed that the protein of SEQ ID NO:268 is a splicing variant of XAGE-1, a CT antigen overexpressed in Ewing sarcoma.

Accordingly, the present invention includes the use of the protein of SEQ ID NO:268, fragments comprising at least 5, 8, 10, 12, 15, 20, 25, 30, 35, 40, 50, 60, 75, 100, 150, or 200 consecutive amino acids thereof, or fragments having a desired biological activity to treat or ameliorate a condition, such as those listed above, associated with over or under expression of the protein of SEQ ID NO:268. In such embodiments, the protein of SEQ ID NO:268, or a fragment thereof, is administered to an individual in whom it is desired to increase or decrease any of the activity of the protein of SEQ ID NO:268. The protein of SEQ ID NO:268 or fragment thereof may be administered directly to the individual or, alternatively, a nucleic acid encoding the protein of SEQ ID NO:268 or a fragment thereof may be administered to the individual. Alternatively, an agent which increases the activity of the protein of SEQ ID NO:268 may be administered to the individual. Such agents may be identified by contacting the protein of SEQ ID NO:268 or a cell or preparation containing the protein of SEQ ID NO:268 with a test agent and assaying whether the test agent increases the activity of the protein. For example, the test agent may be a chemical compound or a polypeptide or peptide.

Alternatively, the activity of the protein of SEQ ID NO:268 may be decreased by administering an agent which interferes with such activity to an individual. Agents which interfere with the activity of the protein of SEQ ID NO:268 may be identified by contacting the protein of SEQ ID NO:268 or a cell or preparation containing the protein of SEQ ID NO:268 with a test agent and assaying whether the test agent decreases the activity of the protein. For example, the agent may be a chemical compound, a polypeptide or peptide, an antibody, or a nucleic acid such as an antisense nucleic acid or a triple helix-forming nucleic acid.

In one embodiment, the invention relates to methods and compositions using the protein of the invention or part thereof as a marker protein to selectively identify tissues, preferably testis and lungs, or to distinguish between two or more possible sources of a tissue sample on the basis of the

level of the protein of SEQ ID NO:268 in the sample. For example, the protein of SEQ ID NO:268 or fragments thereof may be used to generate antibodies using any techniques known to those skilled in the art, including those described therein. Such tissue-specific antibodies may then be used to identify tissues of unknown origin, for example, forensic samples, differentiated tumor tissue that has metastasized to foreign bodily sites, or to differentiate different tissue types in a tissue cross-section using immunochemistry. In such methods a tissue sample is contacted with the antibody, which may be detectably labeled, under conditions which facilitate antibody binding. The level of antibody binding to the test sample is measured and compared to the level of binding to control cells from testis or lungs or tissues other than testis or lungs to determine whether the test sample is from testis or lungs. Alternatively, the level of the protein of SEQ ID NO:268 in a test sample may be measured by determining the level of RNA encoding the protein of SEQ ID NO:268 in the test sample. RNA levels may be measured using nucleic acid arrays or using techniques such as in situ hybridization, Northern blots, dot blots or other techniques familiar to those skilled in the art. If desired, an amplification reaction, such as a PCR reaction, may be performed on the nucleic acid sample prior to analysis. The level of RNA in the test sample is compared to RNA levels in control cells from testis or lungs or tissues other than testis or lungs to determine whether the test sample is from testis or lungs.

In another embodiment, antibodies to the protein of the invention or part thereof may be used for detection, enrichment, or purification of cells expressing the protein of SEQ ID NO:268, including Ewing sarcoma cells, rhabdomyosarcoma cells, breast cancer cells and germ cell tumor cells using methods known to those skilled in the art. For example, an antibody against the protein of SEQ ID NO:268 or a fragment thereof may be fixed to a solid support, such as a chromatography matrix. A preparation containing cells expressing the protein of SEQ ID NO:268 is placed in contact with the antibody under conditions which facilitate binding to the antibody. The support is washed and then the cells are released from the support by contacting the support with agents which cause the cells to dissociate from the antibody.

In another embodiment of the present invention, the protein of SEQ ID NO:268 or a fragment thereof thereof may be used to diagnose disorders associated with altered expression of the protein of SEQ ID NO:268. In some embodiments, the protein of SEQ ID NO:268 or fragments thereof may be used to diagnose Ewing sarcoma, rhabdomyosarcoma, breast cancer or germ cell tumors. In such techniques, the level of the protein of SEQ ID NO:268 in an ill individual is measured using techniques such as those described herein. The level of the protein of SEQ ID NO:268 in the ill individual is compared to the level in normal individuals. An elevated level or decreased level of the protein of SEQ ID NO:268 relative to normal individuals suggests that the ill individual is suffering from a defect in intercellular communication or secretion.

Another embodiment of the invention relates to compositions and methods using the protein of SEQ ID NO:268 or a fragment thereof as possible targets for vaccine-based therapies of cancer,

including Ewing sarcoma, rhabdomyosarcoma, breast cancer or germ cell tumors. In such embodiments, an antibody against the protein of SEQ ID NO:268 or a fragment thereof is administered to an individual suffering from cancer in an amount sufficient to ameliorate or eliminate the cancer.

5 Protein of SEQ ID NO:399 (internal designation (160-40-1-0-H4-CS))

The protein of SEQ ID NO:399 is encoded by the cDNA of SEQ ID NO:158. Accordingly, it will be appreciated that all characteristics and uses of the polypeptide of SEQ ID NO:399 described throughout the present application also pertain to the polypeptide encoded by the human cDNA of clone 160-40-1-0-H4-CS. In addition, it will be appreciated that all characteristics and
10 uses of the nucleic acid of SEQ ID NO:158 described throughout the present application also pertain to the human cDNA of clone 160-40-1-0-H4-CS. The protein of the invention is found to be expressed in testis and lungs. It is over represented in fetal brain.

The protein of SEQ ID NO:399 encoded by the cDNA of SEQ ID NO:158 is homologous to proteins of the Phosphatic Acid Phosphatase type 2 (PAP2) superfamily (Stukey J. and Carman
15 G.M., Protein Sci 1997;6 :469-472, the disclosure of which is incorporated herein by reference in its entirety). Three variants of human PAP, i.e. PAP-alpha 2 (W79285) and its alternatively spliced form PAP-alpha 1 (W79284), PAP-beta (W79286) and PAP-gamma (W79287) have been identified. The protein of SEQ ID NO:399 displays a pfam characteristic domain of the PAP2 superfamily from positions 19 to 175. Accordingly, one embodiment of the present invention is a
20 polypeptide comprising amino acid residues 19 to 175 of SEQ ID NO:399. Four membrane spanning domains are predicted from amino acid positions 17 to, 47 to 67, 108 to 128, and 141 to 161. Accordingly, another embodiment of the present invention is a polypeptide comprising one or more of the foregoing membrane spanning domains.

Phosphatidic acid phosphatase (PAP) (also referred to as phosphatidate phosphohydrolase)
25 is known to be an important enzyme for glycerolipid biosynthesis. In particular, PAP catalyzes the conversion of phosphatidic acid (PA) into diacylglycerol (DAG). PA and DAG are lipids involved in signal transduction and in structural membrane-lipid biosynthesis in cells, thus they represent an important regulatory point in eukariotic phospholipid metabolism. DAG is a well-studied lipid second messenger which is essential for the activation of protein kinase C (Kent; Anal. Rev.
30 Biochem. ; 64 : 315-343;1995; whereas PA itself is also a lipid messenger implicated in various signaling pathways such as NADPH oxidase activation and calcium mobilization (English; Cell Signal.; 8:341-347 ;1996, the disclosure of which is incorporated herein by reference in its entirety). The regulation of PAP activity can therefore affect the balance of divergent signaling processes that the cell receives in terms of PA and DAG (Brindley et al.; Chem.Phys. Lipids 80:45-57 ; 1996, the
35 disclosure of which is incorporated herein by reference in its entirety).

PAP exists in at least two isoforms, one of which (PAP1) is presumed to be cytosolic and membrane associated and the other (PAP2) to be an integral membrane protein (Leung D.W., Tompkins C.K., White T. ; DNA Cell Biol.17 : 377-385 (1998)). The protein of the invention has 180 amino-acids and four predicted membrane-spanning segments, so is presumed to be an integral
5 membrane protein.

The protein of SEQ ID NO:399 is encoded by a cDNA that has homology to many forms of alternative splicing of PAP2 genes. For example, the protein of SEQ ID NO:399 has 29% homology with human phosphatidic acid phosphohydrolase type-2C protein. The protein of SEQ ID NO:399 also has 40% homology with human phosphatidic acid phosphatase 2B protein. In
10 addition, the protein of SEQ ID NO:399 has 33% homology with human type 2 phosphatidic acid phosphatase alpha-2 protein. PAP2-alpha2 is one of the two isoforms with PAP2-alpha1, presumed to be alternative splice variants from a single gene.

Northern analysis has shown that PAP2-alpha mRNA expression was suppressed in several tumor tissues, indicating that PAP-2 may act as a tumor suppressor. The relationship of PAP and
15 tumor suppression is further evidenced in findings that PAP activity is lower in fibroblast cell lines transformed with either the *ras* or *fps* oncogene than in the parental rat1 cell line (Brindley et al ; Chem. Phys. Lipids 80 : 45-57 ;1996, the disclosure of which is incorporated herein by reference in its entirety). As discussed above, a decrease in PAP activity in transformed cells correlates with a concomitant increase in PA concentration. Moreover, elevated PAP activity and lower levels of PA
20 have been observed in contact-inhibited fibroblasts relative to proliferating and transformed fibroblasts (Brindley et al ; Chem. Phys. Lipids 80: 45-57;1996, the disclosure of which is incorporated herein by reference in its entirety). Therefore, the protein of SEQ ID NO:399 or fragments thereof may be used to decrease cell division and as such can provide a useful tool in treating cancer. Subsequent analysis of colon tumor tissue derived from four donors confirmed
25 lower expression of PAP2-alpha than in matching normal colon tissue. Considering these data and previous demonstrations that certain transformed cell lines have lower PAP activity, human PAP cDNAs may be used for gene therapy for certain tumors (Leung D.W., Tompkins C.K., White T. ; DNA Cell Biol.17 : 377-385 (1998), the disclosure of which is incorporated herein by reference in its entirety). Accordingly, one embodiment of the present invention is the use of the protein of
30 SEQ ID NO:399 or a fragment thereof as a tumor suppressor. For example, a nucleic acid expressing the protein of SEQ ID NO:399 or a fragment thereof may be introduced into an individual suffering from cancer in order to ameliorate or eliminate the cancer. In fact, nucleic acids encoding human phosphatidic acid phosphatases have been used to regulate levels of lipid cellular mediators and in gene therapy of e.g. cancer (PCT publication WO98/46730, the disclosure
35 of which is incorporated herein by reference in its entirety).

In another embodiment of the present invention, the protein of SEQ ID NO:399 or a fragment thereof can be used to control the balance of lipid mediators of cellular activation and

signal transduction. The protein of the invention has 33% homology with human phosphatidic acid phosphatase 2A protein. PAP2A is an integral membrane glycoprotein at the cell surface that plays an active role in the hydrolysis and uptake of lipids from the extracellular space (Roberts RZ, Morris AJ; Biochim Biophys Acta 2000 Aug 24;1487(1):33-49, the disclosure of which is
5 incorporated herein by reference in its entirety). Accordingly, the level or activity of the protein of SEQ ID NO:399 may be modulated to influence the rate or extent of hydrolysis and uptake of lipids from the extracellular space using methods such as those described herein.

In another embodiment of the present invention, the protein of SEQ ID NO:399 can be used to counterbalance the inflammatory response. PA has been implicated in cytokine induced
10 inflammatory responses (Bursten et al; Circ. Shok 44: 14-29, 1994; Abraham et al; J. Exp. Med. 181: 569-575, 1995; Rice et al; PNAS 91: 3857-3861, 1994; Leung et al; PNAS 92: 4813-4817, 1995, the disclosures of which are incorporated herein by reference in their entireties) and the modulation of numerous protein kinases involved in signal transduction (English et al ; Chem. Phys. Lipids 80: 117-132, 1996, the disclosure of which is incorporated herein by reference in its
15 entirety). In addition, a nucleic acid encoding the protein of SEQ ID NO:399 or a fragment thereof may be used to counterbalance the inflammatory response from cytokine stimulation through degradation of excess amount of PA in cells or to treat or ameliorate inflammatory diseases.

The gene encoding the protein of SEQ ID NO:399 or a fragment thereof can also be used in gene therapy for the treatment of obesity associated with diabetes. PAP activity is decreased in the
20 livers and hearts of the grossly obese and insulin resistant JCR:LA corpulent rat compared to the control lean phenotype (Brindley et al ; Chem. Phys. Lipids 80 : 45-57 ;1996, the disclosure of which is incorporated herein by reference in its entirety). The protein of the invention therefore can provide an important tool for the treatment of obesity associated with diabetes.

Accordingly, the present invention includes the use of the protein of SEQ ID NO:399 ,
25 fragments comprising at least 5, 8, 10, 12, 15, 20, 25, 30, 35, 40, 50, 60, 75, 100, 150, or 200 consecutive amino acids thereof, or fragments having a desired biological activity to treat or ameliorate a condition, such as those listed above, in an individual. In such embodiments, the protein of SEQ ID NO:399 , or a fragment thereof, is administered to an individual in whom it is desired to increase or decrease any of the activities of the protein of SEQ ID NO:399, including
30 glycerolipid biosynthesis, conversion of phasphatidic acid into diacylglycerol, signal transduction, membrane-lipid biosynthesis, activation of protein kinase C, NADPH oxidase activation, calcium mobilization, cell division, production of diacylglycerol, monoacylglycerol, ceramide or sphingosine, modulation of the inflammatory response or dephosphorylation of a substrate such as lysophasphatidic acid, ceramide 1-phosphate, or sphingosine 1-phosphate, or treatment or
35 amelioration of obesity associated with diabetes. The protein of SEQ ID NO:399 or fragment thereof may be administered directly to the individual or, alternatively, a nucleic acid encoding the protein of SEQ ID NO:399 or a fragment thereof may be administered to the individual.

Alternatively, an agent which increases the activity of the protein of SEQ ID NO:399 may be administered to the individual. Such agents may be identified by contacting the protein of SEQ ID NO:399 or a cell or preparation containing the protein of SEQ ID NO:399 with a test agent and assaying whether the test agent increases the activity of the protein. For example, the test agent
5 may be a chemical compound or a polypeptide or peptide.

Alternatively, the activity of the protein of SEQ ID NO:399 may be decreased by administering an agent which interferes with such activity to an individual. Agents which interfere with the activity of the protein of SEQ ID NO:399 may be identified by contacting the protein of SEQ ID NO:399 or a cell or preparation containing the protein of SEQ ID NO:399 with a test agent
10 and assaying whether the test agent decreases the activity of the protein. For example, the agent may be a chemical compound, a polypeptide or peptide, an antibody, or a nucleic acid such as an antisense nucleic acid or a triple helix-forming nucleic acid.

In one embodiment, the invention relates to methods and compositions using the protein of the invention or part thereof as a marker protein to selectively identify tissues, preferably brain, or
15 to distinguish between two or more possible sources of a tissue sample on the basis of the level of the protein of SEQ ID NO:399 in the sample. For example, the protein of SEQ ID NO:399 or fragments thereof may be used to generate antibodies using any techniques known to those skilled in the art, including those described therein. Such tissue-specific antibodies may then be used to identify tissues of unknown origin, for example, forensic samples, differentiated tumor tissue that
20 has metastasized to foreign bodily sites, or to differentiate different tissue types in a tissue cross-section using immunochemistry. In such methods a tissue sample is contacted with the antibody, which may be detectably labeled, under conditions which facilitate antibody binding. The level of antibody binding to the test sample is measured and compared to the level of binding to control cells from brain or tissues other than brain to determine whether the test sample is from brain.
25 Alternatively, the level of the protein of SEQ ID NO:399 in a test sample may be measured by determining the level of RNA encoding the protein of SEQ ID NO:399 in the test sample. RNA levels may be measured using nucleic acid arrays or using techniques such as in situ hybridization, Northern blots, dot blots or other techniques familiar to those skilled in the art. If desired, an amplification reaction, such as a PCR reaction, may be performed on the nucleic acid sample prior
30 to analysis. The level of RNA in the test sample is compared to RNA levels in control cells from brain or tissues other than brain to determine whether the test sample is from brain.

In another embodiment, antibodies to the protein of the invention or part thereof may be used for detection, enrichment, or purification of cells expressing the protein of SEQ ID NO:399, including using methods known to those skilled in the art. For example, an antibody against the
35 protein of SEQ ID NO:399 or a fragment thereof may be fixed to a solid support, such as a chromatography matrix. A preparation containing cells expressing the protein of SEQ ID NO:399 is placed in contact with the antibody under conditions which facilitate binding to the antibody. The

support is washed and then the cells are released from the support by contacting the support with agents which cause the cells to dissociate from the antibody.

In another embodiment of the present invention, the protein of SEQ ID NO:399 or a fragment thereof thereof may be used to diagnose disorders associated with altered expression of the protein of SEQ ID NO:399. In some embodiments, the protein of SEQ ID NO:399 or fragments thereof may be used to diagnose cancer. In such techniques, the level of the protein of SEQ ID NO:399 in an ill individual is measured using techniques such as those described herein. The level of the protein of SEQ ID NO:399 in the ill individual is compared to the level in normal individuals. An elevated level or decreased level of the protein of SEQ ID NO:399 relative to normal individuals suggests that the ill individual may suffer from cancer or be predisposed to getting cancer in the future.

In another embodiment, the present invention relates to methods of preparing a PAP protein of SEQ ID NO:399 comprising the steps of (i) transforming a host cell with an expression vector comprising a polynucleotide encoding SEQ ID NO:399, (ii) culturing the transformed host cells which express the protein and (iii) isolating the protein. The present invention also relates to a method of dephosphorylating a substrate comprising contacting the substrate with an effective amount of isolated protein of SEQ ID NO:399 or a fragment thereof such that the protein catalyzes the dephosphorylation of the substrate. It is further provided that this method occurs *in vitro*, and comprises a step of isolating the dephosphorylated substrate. Additionally, the method can occur *in vivo*, and is effected by the administration of the protein of the invention (or part of it) to a mammal in need thereof.

Protein of SEQ ID NOs:258 and 262 (internal designations 110-007-1-0-C7-CS, 116-055-1-0-A3-CS):

The protein of SEQ ID NO:258 is encoded by the cDNA of SEQ ID NO:17. Accordingly, it will be appreciated that all characteristics and uses of the polypeptide of SEQ ID NO:258 described throughout the present application also pertain to the polypeptide encoded by the human cDNA of clone 110-007-1-0-C7-CS. In addition, it will be appreciated that all characteristics and uses of the nucleic acid of SEQ ID NO:17 described throughout the present application also pertain to the human cDNA of clone 110-007-1-0-C7-CS. The protein of SEQ ID NO:258 shows homologies to two high affinity IgE receptor-like proteins (IGER) with GENESEQP accession numbers W96745 and W41056, the disclosures of which are incorporated herein by reference in their entireties. The protein of SEQ ID NO:258 is expressed in liver and testis. The protein of SEQ ID NO:262, encoded by SEQ ID NO:21, is a variant of the protein of SEQ ID NO:258 and shares all the potential uses and functions described herein. This protein and cDNA share all of the characteristics and uses of the clone, and product thereof, 116-055-1-0-A3-CS).

Like the two high affinity IgE receptor-like proteins, the protein of the invention contains four transmembrane spanning domains of 20 amino acids, between amino acids 53-73, 79-99, 121-141 and 158-178, respectively. The protein of SEQ ID NO:258 crosses the plasma membrane four times forming two small extracellular loops and has both the N- or C- terminals in the cytoplasm.

5 Moreover, the protein of the invention contains a signal peptide (cleavage site at position 21).

The predicted structure of the protein of SEQ ID NO:258 demonstrates the relationship of this protein to FcεRIβ and CDC20 antigen and provides evidence for a family of 4-transmembrane spanning proteins. The conservation of amino acids between all three proteins is highest in the four transmembrane domains. While greater divergence exists in the hydrophilic amino and carboxyl
10 termini, several amino acids within these regions are conserved such as the presence of 4 prolines in the amino terminus of all three proteins. In addition, two cysteine residues (position 147 and 156) are present in the second extracellular domain between TM3 and TM4. This suggests that inter- or intra-molecular di-sulfite bonds in this domain are present in all three proteins.

FcεRI, is part of a tetrameric receptor complex consisting of an α chain, a β chain and two γ
15 chains (Kinet et al. Proc Natl. Acad. Sci. USA, 15: 6483-6487 (1988), the disclosure of which is incorporated herein by reference in its entirety). Together, they mediate interaction with IgE-bound antigens leading to dramatic cellular responses, such as the massive degranulations of mast cells. The β subunit is a 4-transmembrane protein with both the amino and carboxyl termini residing in the cytoplasm.

20 Chromosome mapping localized cDNA of SEQ ID NO:17 to chromosome 11q12, the location of the CD20 gene. However, the murine FcεRIβ and Ly-44 (the murine equivalent of CD20) are both located in the same position in mouse in chromosome 19 (Teder, T.F. et al., J. Immunol. 141:4388-4394 (1988), Clark E.A. and Lane, J.L. Annu. Rev. Immunol. 9:97-127 (1991), the disclosures of which are incorporated herein by reference in their entireties). Therefore, the
25 three genes are believed to have been originated and evolved from the same locus, further supporting the proposition that they are members of the same family of related proteins.

On the basis of the foregoing information, it is believed that the protein of SEQ ID NO:258 is a high affinity immunoglobulin E receptor-like protein.

Atopic diseases, which include allergy, asthma, atopic dermatitis (or eczema) and allergic
30 rhinitis are generally defined as a disorder of Immunoglobulin E (IgE) responses to common antigens, such as pollen or house dust mites. It is frequently detected by either elevated total serum IgE levels, antigen specific IgE response or positive skin tests to common allergens. In principle, atopy can result from dysregulation of any part of the pathway which begins with antigen exposure and IgE response to the interaction of IgE with its receptor on mast cell, the high affinity Fc
35 receptor FcεRI, and the subsequent cellular activation mediated by that ligand-receptor engagement (Ravetch, Nature Genetics, 7: 117-118 (1994), the disclosure of which is incorporated herein by reference in its entirety).

Accordingly, the protein of SEQ ID NO:258 or fragments comprising at least 5, 8, 10, 12, 15, 20, 25, 30, 35, 40, 50, 60, 75, 100, 150, or 200 consecutive amino acids thereof, or fragments having a desired biological activity may administered to an individual in whom it is desired to increase or decrease the activity of the protein of SEQ ID NO:258. In particular, the protein of SEQ ID NO:258 or fragment thereof may be administered to an individual in whom it is desired to regulate the extent of the IgE response. In such methods, the protein of SEQ ID NO:258 or fragment thereof may be administered directly to the individual or, alternatively, a nucleic acid encoding the protein of SEQ ID NO:258 or a fragment thereof may be administered to the individual. Alternatively, an agent which increases the activity of the protein of SEQ ID NO:258 may be administered to the individual. Such agents may be identified by contacting the protein of SEQ ID NO:258 or a cell or preparation containing the protein of SEQ ID NO:258 with a test agent and assaying whether the test agent increases the activity of the protein. For example, the test agent may be a chemical compound or a polypeptide or peptide.

The protein of SEQ ID NO:258 or fragments thereof may also be used to identify genes or polypeptides that may play a rôle in IgE responses or atopic disease. In particular, binding partners for the protein of SEQ ID NO:258 or the genes encoding such binding partners may be identified using a variety of techniques familiar to those skilled in the art, including the techniques described herein.

The protein of SEQ ID NO:258 or the polynucleotide encoding the protein of SEQ ID NO:258 may also be used to diagnose hereditary atopy. In particular, the level of the protein of SEQ ID NO:258 may be determined in a test individual using methods such as those described herein and compared to the levels of normal individuals and individuals suffering from hereditary atopy to determine whether the test individual is suffering from or at risk of suffering hereditary atopy. Alternatively, a nucleic acid sample may be obtained from a test individual and analyzed to determine whether it contains a level of RNA encoding the protein of SEQ ID NO:258 which is associated with hereditary atopy or a mutation in the gene encoding the protein of SEQ ID NO:258 which is associated with hereditary atopy. For example, a nucleic acid sample from the test individual may be contacted with a nucleic acid probe comprising the nucleic acid encoding the protein of SEQ ID NO:258 or a fragment thereof to determine the RNA level or whether the individual has a mutation associated with hereditary atopy. The probe may be either DNA, including cDNA or genomic DNA, or the probe may be RNA. Any of the methods familiar to those skilled in the art may be used in these diagnostic methods, including the methods described herein. For example, the presence of a mutation associated with hereditary atopy can be determined using methods generally known in the art, such as but not limited to PCR, sequencing or mini sequencing as described in the method of Yamamoto et al. (Biochem. Biophys. Res. Comm., 182:507 (1992), the disclosure of which is incorporated by reference herein in its entirety).

The protein of SEQ ID NO:258 can also be used to characterize the induction of expression of FcεRI and the particular function of FcεRIβ. As such, the protein of the invention can be useful in, for example, the design of drugs that block or inhibit induction or activity of FcεRI, thereby treating atopic diseases. In particular, test agents which block or inhibit induction or activity may
5 be identified using the methods described herein.

In an other embodiment, the protein of SEQ ID NO:258 can be employed in the preparation of antibodies, such as monoclonal antibodies, according to methods known in the art, including those described herein. The antibodies can be used to block or mimic ligand binding to the receptor comprising the protein of the invention or other receptors, such as but not limited to FcεRI. The
10 antibodies can also be used to isolate the protein of SEQ ID NO:258 or cells which express the protein of SEQ ID NO:258 using methods such as those described herein. For example, the antibodies may be used to measure the presence of cells containing the protein of SEQ ID NO:258 (including but not limited to hematopoietic cells) in a sample. For example, the method comprises contacting the sample with the antibody under conditions sufficient for the antibody to bind to the
15 protein of SEQ ID NO:258 and detecting the presence of bound antibody using methods known in the art, including those described herein.

In one embodiment, the invention relates to methods and compositions using the protein of the invention or part thereof as a marker protein to selectively identify tissues, preferably liver and testis, or to distinguish between two or more possible sources of a tissue sample on the basis of the
20 level of the protein of SEQ ID NO:258 in the sample. For example, the protein of SEQ ID NO:258 or fragments thereof may be used to generate antibodies using any techniques known to those skilled in the art, including those described therein. Such tissue-specific antibodies may then be used to identify tissues of unknown origin, for example, forensic samples, differentiated tumor tissue that has metastasized to foreign bodily sites, or to differentiate different tissue types in a tissue
25 cross-section using immunochemistry. In such methods a tissue sample is contacted with the antibody, which may be detectably labeled, under conditions which facilitate antibody binding. The level of antibody binding to the test sample is measured and compared to the level of binding to control cells from liver or testis or tissues other than liver or testis to determine whether the test sample is from liver or testis. Alternatively, the level of the protein of SEQ ID NO:258 in a test
30 sample may be measured by determining the level of RNA encoding the protein of SEQ ID NO:258 in the test sample. RNA levels may be measured using nucleic acid arrays or using techniques such as in situ hybridization, Northern blots, dot blots or other techniques familiar to those skilled in the art. If desired, an amplification reaction, such as a PCR reaction, may be performed on the nucleic acid sample prior to analysis. The level of RNA in the test sample is compared to RNA levels in
35 control cells from liver or testis or tissues other than liver or testis to determine whether the test sample is from liver or testis.

Protein of SEQ ID NO:279 (internal designation 160-58-3-0-H3-CS)

The protein of SEQ ID NO:279 is encoded by the cDNA of SEQ ID NO:38. Accordingly, it will be appreciated that all characteristics and uses of the polypeptide of SEQ ID NO:279 described throughout the present application also pertain to the polypeptide encoded by a nucleic acid included in clone 160-58-3-0-H3-CS. In addition, it will be appreciated that all characteristics and uses of the nucleic acid of SEQ ID NO:38 described throughout the present application also pertain to the nucleic acid included in clone 160-58-3-0-H3-CS.

The protein of SEQ ID NO:279 is encoded by a nucleic acid of 1330 nucleotides with an ORF between nt 198 to 998 yielding a 267 amino acid protein. The protein is a polymorphic variant of the sequence (SP:P01210) for proenkephalin A precursor (contains Met- and Leu- enkephalins). It has a signal peptide spanning 24 amino acid and 2 signature motifs for vertebrate endogenous opioid neuropeptides and endogenous opioid neuropeptide precursors. PSORT gives a predicted extracellular localization, including the cell wall (66.7%). The protein of SEQ ID NO:279 is primarily distributed the fetal brain, although expression in other tissues has also been shown (see below). The polymorphic variation is found at amino acid position 75 (E→D, a conservative amino acid change). After signal peptide cleavage (amino acid 47 to 267; 220 amino acid), the protein still contains the polymorphic variation, which is now at amino acid position 29. This does not change any of the sequence of the different enkephalins that result after cleavage of this precursor protein. In addition, the polymorphism is 25 amino acids away from the first cleavage site on the amino terminal side. This is unlikely to change the secondary structure of the actual cleavage site.

PCT publication WO9606863-A1, the disclosure of which is incorporated herein by reference in its entirety, discloses a protein having high homology with the protein of SEQ ID NO:279. Accordingly, the protein of SEQ ID NO:279 is believed to be an enkephalin. Met- and Leu- enkephalins compete with and mimic the effects of opiate drugs. These two pentapeptides with potent opiate agonist activity in bioassay systems were originally identified by Hughes et al (Nature, 258, 577-580, 1975). The natural ligands for opiate receptors, which differ only in their COOH terminal amino acid, were named Met- and Leu-enkephalin to reflect their origin from the brain. Peptides containing these sequences are termed opiate or opioid peptides. Enkephalins are widely distributed throughout the central nervous system in enkephalinergic neuronal networks, and also exist in the peripheral nervous system, for example in autonomic ganglia. Data, largely circumstantial, suggest wide-ranging involvement of endogenous opioids for example in the modulation of pain perception, in mood and behaviour, learning and memory, responses to stress, diverse neuroendocrine functions, immune regulation and cardiovascular and respiratory function.

Met-enkephalin enhances the immune reaction in patients with cancer or AIDS. It can bind opioid receptors present in peripheral inflamed tissues to mediate an analgesic effect.

After exogenous administration of the different enkephalins, several immunologic functions are affected, including antibody production, NK cell activity against tumors and viral infections,

macrophage and polymorphonuclear leukocyte functions, graft rejections, and mitogen-stimulated lymphocyte proliferation. The effects can be bi-directional, where low concentrations enhance, and high concentrations inhibit the same immune function. Thus, enkephalins are modulators of immune reactions.

5 These opioid neuropeptides are released by post-translational proteolytic processing of precursor proteins. These multivalent precursor proteins (polyprotein) consist of a signal sequence followed by a conserved region of about 50 residues, a variable length region and the sequence of the various neuropeptides. The preproenkephalin A (gene PENK) is processed to produce the following peptides which include Met-enkephalin (6 copies, 2 of which are extended) and Leu-
10 enkephalin:

Signal peptide 1-24

Peptide 100-104 Met-enkephalin 1

Peptide 107-111 Met-enkephalin 2

Peptide 136-140 Met-enkephalin 3

15 Peptide 186-193 Met-enkephalin-arg-gly-leu

Peptide 210-214 Met-enkephalin 4

Peptide 230-234 Leu-enkephalin

Peptide 261-267 Met-enkephalin-arg-phe

The conserved region in the N-termini of these precursors contains six cysteines that are
20 probably involved in disulfide bonds. This region could also be important for the processing of the neuropeptides.

The precursor protein does have the potential to be differentially cleaved into multiple extended enkephalin and non-enkephalin-containing peptides, the functions of which are largely unknown; however, in some cases it has been shown that extended enkephalin-containing peptides
25 have enhanced opiate activity. Another peptide, enkelytin, is produced that exhibits anti-bacterial activity (see below).

There is a growing body of evidence that proenkephalin exists largely independently of free enkephalin peptides in a number of tissues and cell types including astrocytes (Melner et al, EMBO J, 9, 791-796, 1990; Spruce et al, EMBO J 9, 1787-1795, 1990, the disclosures of which are
30 incorporated herein by reference in their entirety), and is released from these cells in an unprocessed form (Batter et al, Brain Res. 563, 28-32, 1991, the disclosure of which is incorporated herein by reference in its entirety). There is evidence in some cases that processing enzymes are co-released along with the unprocessed precursor which suggests that extracellular cleavage may occur (Vilijn et al, J. Neurochem. 53, 1487-1493, 1989). Even if biological activity is signalled through
35 binding of the small peptide products to cell surface receptors, the regulation of this activity may be mediated through the precursor, and it is also possible that the unprocessed precursor has an additional intracellular role of its own.

This protein was originally described to be present in various brain regions, most notably in the striatum as well as in neuroendocrine tissues, the pituitary and adrenal gland. It is also expressed in a variety of immune cells, including ConA-stimulated CD4 T lymphocytes, CD4 thymocytes, B lymphocytes, as well as T cell lines, macrophages and mast cells. Expression has been reported in the reproductive system, heart and many developing tissues during gestation and early postnatal period. Because of this, it has been postulated that these peptides play a role in cell or tissue growth and differentiation. For example, endogenous enkephalins induced in thymocytes modulate their own expression and function to inhibit the proliferation of activated thymocytes.

Enkephalin peptides are abundant in adrenal medulla and can be released by neurotransmitters specific for that tissue. Enkephalins have also been found to be abundant in human pheochromocytoma, a tumour derived from the adrenal medulla. The RNA from this tumour contains a high level of enkephalin mRNA sequences as demonstrated by cell-free translation studies.

Enkephalins function as opiate receptors are classified as delta, kappa and mu. A study by Lord et al (Nature, 267, 495-499, 1977) compared the activity of morphine and enkephalins in bioassay systems, and found that enkephalins bound predominantly to delta receptors. Subsequent studies have revealed homology of these receptors to other receptor families, including the immunoglobulin superfamily member OBCAM (Schofield et al, EMBO J 8, 489-495, 1989, the disclosure of which is incorporated herein by reference in its entirety) and somatostatin receptors (PCT publication WO96/06863, the disclosure of which is incorporated herein by reference in its entirety). This would explain the reported opioid binding properties of the former. Because of the latter's homology to opiate receptors, it would also be expected to bind opioid receptor ligands. The recognition of opioid peptides by other non-opiate related receptors implies that these peptides may exert other as yet unknown functions.

Enkephalins are also involved in apoptosis. Apoptosis is the morphologically distinct process of controlled cell death which balances the process of cell production by mitosis. A molecular connection between control of cell production and cell elimination has now been established, including the roles of c-myc and p53 in the pathways mediating apoptotic cell death. It has been proposed that all mammalian cells may be programmed to die by default in the absence of continuous signalling from neighboring cells. However, the acquisition of a survival advantage which prevents a single cell from activating its suicide program in response to levels of genetic damage associated with common environmental insults could theoretically be an initiating event in oncogenesis since it would favor the persistence of potentially tumorigenic mutations. Alternatively, inappropriate activation of survival pathways might lead to overriding the intrinsic death program and promote tumorigenesis at early and late stages. A particularly potent oncogenic pathway would be one which both promoted and tolerated genetic damage and helped a cell overcome its need for extracellular survival signals. Approximately 50% of human tumors possess

normal p53 function. Thus, additional pathways or molecules which inappropriately repress apoptosis in human tumours remain to be identified. Opioid-like molecules could be involved in such a pathway.

There are published reports that pathways which include opioid-like molecules participate in regulating the equilibrium between cell death and survival. For example, morphine inhibits cell survival in the developing cerebellum (Hauser et al, Exp. Neurol, 130, 95-105, 1994, the disclosure of which is incorporated herein by reference in its entirety) and induces apoptosis in thymocytes (Fuchs and Pruett, J. Pharmacol. Exp. Ther. 266, 417-423, 1993, the disclosure of which is incorporated herein by reference in its entirety).

In a series of experiments (PCT publication WO 96/06863), it has been found that proenkephalin and/or its proteolytic products act as extracellular and/or cell surface membrane bound factors which modulate cell survival in transformed cells a) upon deprivation of exogenous survival factors, and b) following genotoxic injury and/or stress when exogenous survival factors are non-limiting. The receptor(s) to which these factor(s) bind, which are most likely to exist on the cell surface are related, or possibly identical, to one or more members of the opioid receptor family.

Opioid-like receptor types or subtypes can mediate survival or death; receptor (s) which mediate death appear to be coupled to those which mediate survival. Natural ligands for these receptors are likely to be products of the opioid precursor genes, although natural ligands could include cytokines which mimic their effect. Tumour cells are more sensitive to antagonism of opioid-like receptor-mediated survival, and to stimulation of opioid-like receptor-mediated death, than non-transformed cells. The induction of cell cycle arrest enhances the sensitivity of tumour cells to these manipulations. (Enhanced sensitivity of tumour cells to these manipulations is induced by their synchronisation within the cell cycle.

Cytoplasmic proenkephalin and/or its proteolytic products act as general repressors of apoptosis. Agents which, if coupled to appropriate internalisation agents, would antagonise cytoplasmic proenkephalin would therefore be of use in the induction of apoptosis in non-transformed as well as transformed cells, particularly in combination with sublethal doses of known apoptosis-inducing agents.

The repression of apoptosis mediated through cytoplasmic proenkephalin is activated at high cell density predominantly by nondiffusible factors. Inhibition of proenkephalin or its products as described above would therefore be potentiated if agents were used in combination for example with neutralising antibodies to integrins (such as the antibody 23C6- Bates et al., J. Cell Biol. 125 403-415, 1994) to reduce exogenous survival signaling and simulate low density.

Proenkephalin targeted to the cell nucleus induces apoptotic death, which is inhibited by the overexpression of large T antigen and is at least partly mediated through p53. Tumors which retain wild-type p53 function are therefore a particular target for apoptosis induction by agents which

increase the levels of proenkephalin, or its derivatives, within the nucleus or which mimic the function of nuclear proenkephalin or its derivatives.

Accordingly, the protein of SEQ ID NO:279, fragments thereof, or nucleic acids encoding the protein of SEQ ID NO:279 may be used to modulate a biochemical pathway in which products of opioid peptide precursor genes participate. In some embodiments, antibodies or other agents which reduce the level or activity of the protein of SEQ ID NO:279 or fragments thereof may be used to induce apoptosis in cells. The agents preferably neutralize the protein of SEQ ID NO:279 or its proteolytic derivatives, increase the level of, activate or mimic nuclear proenkephalin, or act as an antagonist to receptors related or identical to the delta and kappa opioid receptors. In some embodiments, the agent may be a neutralizing monoclonal antibody against the protein of SEQ ID NO:279 or a fragment thereof. The agent may also be a fragment or allelic form of one of these antibodies. A cytoplasmic anchor, or a nuclear localization signal may also be included in the agent. In some embodiments, the agent is able to modulate a biochemical pathway in a cell in which products of opioid peptide precursor genes participate in order to induce apoptosis. The agents can be used for the treatment of cancer or for inducing apoptosis in lens cells following a cataract operation. In some embodiments, the agents promote apoptosis of proliferating cells with less, or no, effect on normal mature cell types. The agents may be administered in combination with a genotoxic or cell cycle arrest agent. Alternatively, the agent may be complexed with a chemotherapeutic, irradiation or cell cycle arrest (synchronization agent).

Accordingly, the invention provides a means of inducing apoptosis in cells which comprises modifying a biological pathway of a cell in which a product of an opioid precursor gene participates in such a way that apoptosis is induced. Modification of the pathway is suitably effected by administration of an appropriate agent. In particular, the present invention provides an agent for use in inducing apoptosis in cells, said agent comprising an agent able to neutralise proenkephalin or its proteolytic derivatives; an agent which increases the level of nuclear proenkephalin and/or its derivatives, or which activates or mimics them an agent which acts as an antagonist at receptor(s) related or identical to the delta opioid receptor, or an agent which acts as an agonist at receptor(s) related or identical to the kappa opioid receptor.

A subset of such agents are agents able to neutralise proenkephalin or its proteolytic derivatives, or an agent which acts as an antagonist at receptor(s) related or identical to the delta opioid receptor, or an agent which acts as an agonist at receptor(s) related or identical to the kappa opioid receptor.

In some embodiments, the agent may be administered to the cell surface whereupon the survival effects of extracellular and/or cell surface membrane bound proenkephalin or its proteolytic derivatives is neutralised causing the cell to become apoptotic. Alternatively, an agent able to neutralise proenkephalin or its proteolytic derivatives may be coupled to an internalisation peptide and a cytoplasmic anchor. Such an assembly will remain in the cytoplasm of the cell, antagonising

cytoplasmic proenkephalin and/or its proteolytic products and thus neutralising the apoptosis repressor effect of these molecules.

Enkephalins also have anti-bacterial activity. During processing of the proenkephalin-A, the maturation in the adrenal medullary chromaffin cell starts with the removal of the carboxy-terminal end (proenkephalin-A-derived peptide or PEAP₂₀₉₋₂₃₉) (Y. Goumon, K. Lugardon, B. Kieffer et al. J. Biol. Chem. 273:29847-29856, 1998, the disclosure of which is incorporated herein by reference in its entirety). The peptide enkelytin was identified as corresponding to bisphosphorylated PEAP₂₀₉₋₂₃₇, and possesses antibacterial activity including *Staphylococcus aureus* and other gram-positive bacteria such as *Micrococcus luteus* and *Bacillus megaterium* (0.2-0.4 μ M range). There is no ability to affect gram-negative bacteria (*E. coli* strain D22, D31, 663 and T13773) growth, nor is there any hemolytic activity. The activity of this peptide is specific – shorter versions of the peptide (209-220, 224-237, 230-237, 233-237) or non-phosphorylated PEAP₂₀₉₋₂₃₉ exhibited little to no bacterial growth inhibiting activity.

Bovine periarthritis abscess fluid contains different forms of PEAP (72-237/239; 80-237/239) as identified by immunoreactivity and confirmed by sequence analysis. These peptides have activity against *M. luteus*, but are less active than enkelytin (5 versus 0.2 μ M). These PEAP constitute a pool of precursors which have to be processed, during infection, to provide active enkelytin. Presence of a PEAP at a molecular mass corresponding to that of PEAP₂₀₉₋₂₃₇ was detected as well. PEAPs (PEAP₂₀₂₋₂₃₈ and PEAP₂₀₆₋₂₃₇) have also been detected in wound fluids, including bovine post-caesarean abscess in the subcutaneous lining, and an abscess induced by subcutaneous injection of complete Freund's adjuvant. Therefore, these peptides are present in wound fluids along with other known antibacterial peptides (defensins, batenecins). The concentrations were in a range similar to that found to be active *in vitro* (0.5-1 μ M). The PEAPs have also been detected in secretions from human polymorphonuclear neutrophils.

The PEAP₂₀₉₋₂₃₀ and enkelytin are secreted from cultured chromaffin cells following stimulation. This suggests that these two peptides are co-released with catecholamines in stress situations and may therefore play an important role in defense mechanisms.

Co-release of met-enkephalin and enkelytin would represent a unified neuroimmune protective response to stress situations that may be accompanied with infectious diseases. This would provide a highly beneficial survival strategy at the very beginning of proinflammatory processes. This protein would therefore play an important role in host defense against microbial infections, especially those involving gram positive bacteria. Due to their nonspecific activity on membranes, the antibacterial peptides possess cytotoxic activities and may not only play a role in antimicrobial defense, but also in inflammatory processes, possibly in wound repair.

The protein of SEQ ID NO:279, peptides derived by cleavage thereof or fragments thereof could be used as antibacterial agents in creams/ointments/solutions, presoaked bandages, or dermal-type patches for external applications. Alternatively, the protein of SEQ ID NO:279, peptides

derived by cleavage thereof, or fragments thereof may be used in injections (intravenously, subcutaneously or intra-peritoneally). This is useful for wound repair, burn healing, post-operative recovery management.

Alternatively, the protein of SEQ ID NO:279, peptides derived by cleavage thereof, or
5 fragments thereof, may be incorporated into disinfectant solutions used for cleaning surfaces such as in the the house (kitchen, bathroom) or in the office (desktops, phones, computer keyboards and mouse). Other applications are as additives in mouthwash or handi-popup wipes.

Altered levels of enkephalins may produce psychological disease. Konig et al (Nature, 383, 535-538, 1996, the disclosure of which is incorporated herein by reference in its entirety) used a
10 genetic approach to study the role of the mammalian opioid system. They disrupted the pre-proenkephalin gene using homologous recombination in embryonic stem cells to generate enkephalin-deficient mice. Mutant enk -/- animals are healthy, fertile, and care for their offspring, but display significant behavioral abnormalities. Mice with the enk -/- genotype are more anxious and males display increased offensive aggressiveness. Mutant animals show marked differences
15 from controls in supraspinal, but not in spinal, responses to painful stimuli. These enk -/- mice do however exhibit normal stress-induced analgesia. Therefore, enkephalins modulate responses to painful stimuli. Thus, genetic factors may contribute significantly to the experience of pain. This study clearly indicates the importance of enkephalins in pain perception, anxiety and aggressiveness.

20 Interestingly, the PENK gene is localized on 8q23-q24, the same locus on which are found genes related to epilepsy and spastic paraplegia, disorders related to brain dysfunction.

Accordingly, the protein of SEQ ID NO:279 or fragments thereof may be used for the treatment of psychological disorders, especially those involving distortion in the perception of pain, aggressiveness, or anxiety. This would include drug addiction, different types of phobias, panic
25 attacks, schizophrenia, bi-polar, anorexia nervosa, chronic pain disorders, post-traumatic events. post-operative pain management.

Accordingly, the present invention includes the use of the protein of SEQ ID NO:279, fragments comprising at least 5, 8, 10, 12, 15, 20, 25, 30, 35, 40, 50, 60, 75, 100, 150, or 200 consecutive amino acids thereof, or fragments having a desired biological activity to treat or
30 ameliorate a condition in an individual. For example, the condition may be cancer, a condition resulting from increased or decreased cellular proliferation, bacterial infection, conditions resulting from abnormal immune responses, psychological disease or any of the conditions listed above. In such embodiments, the protein of SEQ ID NO:279, or a fragment thereof, is administered to an individual in whom it is desired to increase or decrease any of the activities of the protein of SEQ
35 ID NO:279. The protein of SEQ ID NO:279 or fragment thereof may be administered directly to the individual or, alternatively, a nucleic acid encoding the protein of SEQ ID NO:279 or a fragment thereof may be administered to the individual. Alternatively, an agent which increases the

activity of the protein of SEQ ID NO:279 may be administered to the individual. Such agents may be identified by contacting the protein of SEQ ID NO:279 or a cell or preparation containing the protein of SEQ ID NO:279 with a test agent and assaying whether the test agent increases the activity of the protein. For example, the test agent may be a chemical compound or a polypeptide
5 or peptide.

Alternatively, the activity of the protein of SEQ ID NO:279 may be decreased by administering an agent which interferes with such activity to an individual. Agents which interfere with the activity of the protein of SEQ ID NO:279 may be identified by contacting the protein of SEQ ID NO:279 or a cell or preparation containing the protein of SEQ ID NO:279 with a test agent
10 and assaying whether the test agent decreases the activity of the protein. For example, the agent may be a chemical compound, a polypeptide or peptide, an antibody, or a nucleic acid such as an antisense nucleic acid or a triple helix-forming nucleic acid.

In one embodiment, the invention relates to methods and compositions using the protein of the invention or part thereof as a marker protein to selectively identify the source of a sample as, for
15 example, fetal brain, or to distinguish between two or more possible sources of a sample on the basis of the level of the protein of SEQ ID NO:279 in the sample. For example, the protein of SEQ ID NO:279 or fragments thereof may be used to generate antibodies using any techniques known to those skilled in the art, including those described therein. Such antibodies may then be used to identify tissues of unknown origin, for example, forensic samples, differentiated tumor tissue that
20 has metastasized to foreign bodily sites, or to differentiate different tissue types in a tissue cross-section using immunochemistry. In such methods a sample is contacted with the antibody, which may be detectably labeled, under conditions which facilitate antibody binding. The level of antibody binding to the test sample is measured and compared to the level of binding to control cells from fetal brain or tissues other than fetal brain to determine whether the test sample is from fetal
25 brain. Alternatively, the level of the protein of SEQ ID NO:279 in a test sample may be measured by determining the level of RNA encoding the protein of SEQ ID NO:279 in the test sample. RNA levels may be measured using nucleic acid arrays or using techniques such as in situ hybridization, Northern blots, dot blots or other techniques familiar to those skilled in the art. If desired, an amplification reaction, such as a PCR reaction, may be performed on the nucleic acid sample prior
30 to analysis. The level of RNA in the test sample is compared to RNA levels in control cells from fetal brain or tissues other than fetal brain to determine whether the test sample is from fetal brain.

In another embodiment, antibodies to the protein of the invention or part thereof may be used for detection, enrichment, or purification of cells expressing the protein of SEQ ID NO:279, including using methods known to those skilled in the art. For example, an antibody against the
35 protein of SEQ ID NO:279 or a fragment thereof may be fixed to a solid support, such as a chromatography matrix. A preparation containing cells expressing the protein of SEQ ID NO:279 is placed in contact with the antibody under conditions which facilitate binding to the antibody. The

support is washed and then the cells are released from the support by contacting the support with agents which cause the cells to dissociate from the antibody.

In another embodiment of the present invention, the protein of SEQ ID NO:279 or a fragment thereof thereof may be used to diagnose disorders associated with altered expression of the protein of SEQ ID NO:279. In such techniques, the level of the protein of SEQ ID NO:279 in an ill individual is measured using techniques such as those described herein. The level of the protein of SEQ ID NO:279 in the ill individual is compared to the level in normal individuals to determine whether the individual has a level of the protein of SEQ ID NO:279 which is associated with disease.

10 Protein of SEQ ID NO: 293 (internal designation 181-16-1-0-G7-CS)

The protein of SEQ ID NO: 293 has a high degree of homology with HSPC163 (Genbank accession number AF161512), the protein encoded by gene no: 93 (PCT/US99/17130) and the human cornichon protein TGAM77. SEQ ID NO: 293 is overexpressed in cancerous prostate, fetal brain and fetal kidney.

15 The gene HSPC163 is one of three hundred cDNAs obtained from CD34+ hematopoietic stem / progenitor cell (HSPC) library (obtained from umbilical cord blood and adult bone marrow). HSPC163 has also been identified in five hematopoietic cell lines: NB4 (granulocytic), HL60 (granulocytic), U937 (monocytic), K562 (erythro-megakaryocytic), and Jurkat (T lymphocytic). These cell lines represent the distinct lineages of hematopoietic cells.

20 The polypeptide of gene no: 93 has been determined to have two transmembrane domains and a short cytoplasmic tail. Based upon these characteristics, it is believed that the protein product of gene no: 93 shares structural similarity to type IIIa membrane proteins. This gene is expressed primarily in activated T-cells and to a lesser extent in endometrial tumor, T cell helper II cells, microvascular endothelial cells, Raji cells treated with cyclohexamide and umbilical vein
25 endothelial cells. The expression pattern of gene no: 93, indicates a role in regulating the proliferation, survival, differentiation, and/or activation of hematopoietic cell lineages, including blood stem cells. The gene product appears to be involved in the regulation of cytokine production, antigen presentation, and other immune processes, suggesting a usefulness in boosting the immune system. The translation product of this gene has high homology to the human TGAM77 and mouse
30 cornichon proteins.

TGAM77 was identified as a gene involved in early phase of T-cell activation in response to alloantigens. Twenty four hours after T-cell allostimulation, RNA expression of TGAM77 is significantly increased. TGAM77 has been designated as a T-cell growth associated molecule. TGAM77 is a human homolog of cornichon (cni) protein of the fruit fly *Drosophila*.

35 Cornichon was demonstrated to be involved in carefully orchestrated signaling events during *Drosophila* oogenesis establishing an asymmetric pattern in the oocyte as a prerequisite for

correct embryogenesis. Cornichon signaling functions in concert with two other proteins. The function of all three genes in an EGF-like signaling pathway appears to direct the formation of a correctly polarized microtubule cytoskeleton, which is thought to be the basis for the correct spatial localization of other signaling molecules essential for oocyte polarization, asymmetric movement
5 of the nucleus, and embryo differentiation.

The subject invention provides the amino acid sequence of SEQ ID NO: 293 and polynucleotide sequences encoding the amino acid sequence of SEQ ID NO: 293. In one embodiment, the polypeptides of SEQ ID NO: 293 are interchanged with the corresponding polypeptides encoded by the human cDNA of clone 181-16-1-0-G7-CS. Also included in the
10 invention are biologically active fragments of SEQ ID NO: 293 and polynucleotide sequences encoding these biologically active fragments. "Biologically active fragments" are defined as those peptide or polypeptide fragments of SEQ ID NO: 293 which have at least one of the biological functions of the full length protein (e.g., the ability to stimulate T-cell proliferation).

The invention also provides variants of SEQ ID NO: 293. These variants have at least
15 about 80%, more preferably at least about 90%, and most preferably at least about 95% amino acid sequence identity to the amino acid sequence of SEQ ID NO: 293. Variants according to the subject invention also have at least one functional or structural characteristic of SEQ ID NO: 293, such as the biological functions described above. The invention also provides biologically active fragments of the variant proteins. Unless otherwise indicated, the methods disclosed herein can be
20 practiced utilizing SEQ ID NO: 293 or variants thereof. Likewise, the methods of the subject invention can be practiced using biologically fragments of SEQ ID NO: 293, or variants of said biologically active fragments.

Because of the redundancy of the genetic code, a variety of different DNA sequences can encode SEQ ID NO: 293. It is well within the skill of a person trained in the art to create these
25 alternative DNA sequences which encode proteins having the same, or essentially the same, amino acid sequence. These variant DNA sequences are, thus, within the scope of the subject invention. As used herein, reference to "essentially the same" sequence refers to sequences that have amino acid substitutions, deletions, additions, or insertions that do not materially affect biological activity. Fragments retaining one or more characteristic biological activity of SEQ ID NO: 293 are also
30 included in this definition.

"Recombinant nucleotide variants" are alternate polynucleotides which encode a particular protein. They can be synthesized, for example, by making use of the "redundancy" in the genetic code. Various codon substitutions, such as the silent changes which produce specific restriction sites or codon usage-specific mutations, can be introduced to optimize cloning into a plasmid or
35 viral vector or expression in a particular prokaryotic or eukaryotic host system, respectively.

SEQ ID NO: 293 protein, and variants thereof, can be used to produce antibodies according to methods well known in the art. The antibodies can be monoclonal or polyclonal. Antibodies can

also be synthesized against fragments of SEQ ID NO: 293 as well as variants of SEQ ID NO: 293 according to known methods. The subject invention also provides antibodies which specifically bind to biologically active fragments of SEQ ID NO: 293 or biologically active fragments of SEQ ID NO: 293 variants.

5 The subject invention also provides for immunoassays which are used to screen for, monitor, or diagnose prostate cancer. Methods of screening for, diagnosing, identifying, or monitoring the course of prostate cancer are well known to those skilled in the art. In this aspect of the invention, immunoassays are provided which contact a biological sample (e.g., blood, serum, tissue, or biopsied tissue sample) with antibodies which specifically bind to SEQ ID NO: 293 ,
10 immunogenic fragments of SEQ ID NO: 293 , or biologically active fragments of SEQ ID NO: 293 . Immunocomplexes formed in the contacting step are then detected using an appropriately labeled detection reagent. The levels of SEQ ID NO: 293 expressed in the tested biological samples are compared to control/normal levels typically observed in the population.

Alternatively, methods which screen for, monitor, or diagnose prostate cancer may be
15 practiced with SEQ ID NO: 293 , or fragments of SEQ ID NO: 293 , as well as nucleic acids encoding SEQ ID NO: 293 , or fragments of SEQ ID NO: 293 . In one embodiment, the polypeptide may be used as a standard/control immunoassays described above. In another embodiment, the nucleic acids encoding SEQ ID NO: 293 , or fragments of SEQ ID NO: 293 are used in hybridization assays, well known to the skilled artisan, to identify biological samples (e.g.,
20 blood, serum, tissue, or biopsied tissue sample) which contain SEQ ID NO: 293 . The levels of SEQ ID NO: 293 expressed in the tested biological samples are compared to control/normal levels typically observed in the population.

In another embodiment, SEQ ID NO: 293 , and polynucleotide sequences encoding the amino acid sequence of SEQ ID NO: 293 can be used to identify or diagnose immune disorders
25 involving activated T-cells using standard hybridization assays.

Another aspect of the invention provides methods of immunostimulating a mammal. In this aspect of the invention, SEQ ID NO: 293 , and/or polynucleotide sequences encoding the amino acid sequence of SEQ ID NO: 293 , are introduced into T-cells according to well known methods. T-cells are, then activated by stimulation with antigen to induce the immune system of the mammal.

30 In another embodiment, autologous T-cells are obtained from an individual. SEQ ID NO: 293 , biologically active fragments thereof, and/or polynucleotide sequences encoding the amino acid sequence, or biologically active fragments, of SEQ ID NO: 293 , are introduced into these autologous T-cells according to well known methods. The T-cells are expanded and reintroduced into the individual from which the T-cells were obtained. See, for example U.S. Patent Nos.
35 5,192,537 and 5,766,920 , hereby incorporated by reference in their entirety.

In another embodiment of the subject invention, polynucleotides and polypeptides encoding SEQ ID NO: 293 , can be used to expand stem cells, committed progenitors of various

blood lineages, and in the differentiation and/or proliferation of various cell types. In this aspect of the invention, polynucleotides and polypeptides encoding SEQ ID NO: 293 are introduced into the cells and the cells cultured. These methods may be practiced according to methods well known to the routineer.

5 Protein of SEQ ID NO:316 (internal designation 188-45-1-0-D9-CS)

The protein of SEQ ID NO:316 is encoded by the cDNA of SEQ ID NO:75. Accordingly, it will be appreciated that all characteristics and uses of the polypeptide of SEQ ID NO:316 described throughout the present application also pertain to the polypeptide encoded by a nucleic acid included in clone 188-45-1-0-D9-CS. In addition, it will be appreciated that all characteristics
10 and uses of the nucleic acid of SEQ ID NO:75 described throughout the present application also pertain to the nucleic acid included in clone 188-45-1-0-D9-CS.

The protein of SEQ ID NO:316 is expressed in brain and contains three membrane-spanning segments located between amino acid positions 6 and 26, 73 and 93, or 139 and 159 and a signal peptide comprising the sequence FAAFCYMLSLVLC/AA. Accordingly, one embodiment
15 of the present invention is a polypeptide comprising one or more of the membrane-spanning segments, and/or the signal peptide.

The protein of SEQ ID NO:316 is a member of the cornichon protein family. It has 48% identity with the *Drosophila melanogaster* cornichon protein as well as 67% identity with the Human Cornichon homolog TGAM77 (Genbank accession No. AF104398, the disclosure of which
20 is incorporated herein by reference in its entirety), 67% identity with hCornichon, a bone marrow secreted protein (PCT publication WO/9933979, the disclosure of which is incorporated herein by reference in its entirety), 67% identity with a human secreted protein encoded by gene 24 (PCT publication WO/9910363, the disclosure of which is incorporated herein by reference in its entirety) and 67% identity with the protein product of the mouse *cni* gene. However, this protein has higher
25 homology, 81% identity, to the mouse cornichon-like protein (Genbank accession No. AB006191, the disclosure of which is incorporated herein by reference in its entirety), which is the product of the mouse *cni* gene. Finally, the protein of SEQ ID NO:316 has a high level of identity with human secreted protein encoded by gene 95 (GSP:Y76218, PCT publication WO/9958660, the disclosure of which is incorporated herein by reference in its entirety) and is likely a polymorphic
30 variant of gene 95. The high degree of sequence conservation between the members of this family indicates that they are under strong selective pressure and are likely involved in important cellular functions.

The *Drosophila cornichon (cni)* gene product is involved in signaling processes necessary for both anterior-posterior and dorsal-ventral pattern formation during *Drosophila* embryogenesis
35 (Cell, 1995, 81:967-978). Mutations in *cornichon* prevent the formation of a correctly polarized microtubule cytoskeleton in the oocyte. *Cni* signaling functions in concert with two other proteins.

Gurken, which is a protein secreted from the oocyte containing a single epidermal growth factor (EGF) motif most similar in structure to vertebrate TGF α , is considered to be the ligand of the *Drosophila* epidermal growth factor receptor (DER) homolog torpedo, which is expressed by the follicular epithelium. The function of all three genes in an EGF-like signaling pathway appears to
5 direct the formation of a correctly polarized microtubule cytoskeleton, which is thought to be the basis for the correct spatial localization of other signaling molecules essential for oocyte polarization, asymmetric movement of the nucleus, and embryo differentiation. TGAM77, one of the human homologs of cornichon, is differently expressed in alloactivated T-cells (Bioch. Biophys. Acta 1999, 1449:203-210, the disclosure of which is incorporated herein by reference in its
10 entirety). Since there is a well-known involvement of the microtubule cytoskeleton in spatial polarization of signaling events in T-cell activation, it is thought that TGAM77 may function in a protein-tyrosine kinase pathway required for the vectorial localization of signaling molecules in T-cell activation.

The protein of SEQ ID NO:316 is found in brain tissue, and gene 95 (GSP:Y76218, PCT
15 publication WO/9958660, the disclosure of which is incorporated herein by reference) is expressed in infant brain tissue, endometrial tumor tissue and frontal cortex tissue. ESTs matching this gene are also found in lung tissue, germ cell tumors and skin melanomas. This is similar to the expression pattern of the murine *cnil* gene, which is found in 6.5-day whole embryos, 11.5-day limb bud, 13.5-day whole embryo, adult lung and brain (Dev. Genes Evol., 1999, 209:120-125, the
20 disclosure of which is incorporated herein by reference in its entirety).

Polynucleotides encoding the protein of SEQ ID NO:316 or fragments thereof and polypeptides comprising the protein of SEQ ID NO:316 or fragments thereof are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, endometrial tumor, and
25 neural and developmental diseases and/ or disorders. Similarly, the protein of SEQ ID NO:316 or fragments thereof and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neural and reproductive organs, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell
30 types (e.g., neural, reproductive, cancerous and wounded tissues) or bodily fluids (e.g. lymph, serum, plasma, urine, amniotic fluid, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in infant brain tissue and adult brain tissue, as well as the homology
35 to cornichon proteins, indicates that polynucleotides encoding the protein of SEQ ID NO:316 or fragments thereof and polypeptides comprising the protein of SEQ ID NO:316 or fragments thereof are useful for detecting and/or treating neural and developmental disorders. The tissue distribution

indicates that these polynucleotides and polypeptides are useful for the detection/treatment of neurodegenerative disease states and behavioural disorders such as Alzheimers Disease, Parkinsons Disease, Huntingtons Disease, Tourette Syndrome, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, panic disorder, learning disabilities, ALS, Psychoses, autism, and
5 altered behaviors, including disorders in feeding, sleep platterns, balance, and perception. In addition, the gene or gene product may also play a role in treatment and/or detection of developmental disorders associated with the developing embyo, or sexually-linked disorders,

Elevated expression of the protein of SEQ ID NO:316 within the brain suggests that it may be involved in neuronal survival, synapse formation, conductance, neural differentiation, etc. Such
10 involvement may impact many processes, such as learing and cognition. Alternatively, the tissue distribution in endometiral tumor tissue, germ cell tumors and skin melanomas indicates that the translation product of this gene is useful for the detection and/or treatment of endometrial tumors and/or reproductive disorders, as well as tumors of other tissues where expression of this gene has been observed. Furthermore, the protein of SEQ ID NO:316 or fragments thereof may also be used
15 to determine biological activity, to raise antibodies, as a tissue marker, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. The protein of SEQ ID NO:316 or fragments thereof, as well as, antibodies directed against the protein may be used as tumor marker and/or immunotherapy targets for the above listed tissues.

20 The gene encoding the protein of SEQ ID NO:316 is thought to reside on chromosome 11. Accordingly, polynucleotides encoding the protein of SEQ ID NO:316 or fragments thereof are useful as a marker in linkage analysis for chromosome 11.

Accordingly, the present invention includes the use of the protein of SEQ ID NO:316 , fragments comprising at least 5, 8, 10, 12, 15, 20, 25, 30, 35, 40, 50, 60, 75, 100, 150, or 200
25 consecutive amino acids thereof, or fragments having a desired biological activity to treat or ameliorate a condition in an individual. For example, the condition may be an abnormality in development, a signaling pathway, microtubule construction, neuronal survival, synapse formation, conductance, neuarl differentiation, or it may be cancer or an abnormality in any of the functions listed above. In such embodiments, the protein of SEQ ID NO:316, or a fragment thereof, is
30 administered to an individual in whom it is desired to increase or decrease any of the activities of the protein of SEQ ID NO:316. The protein of SEQ ID NO:316 or fragment thereof may be administered directly to the individual or, alternatively, a nucleic acid encoding the protein of SEQ ID NO:316 or a fragment thereof may be administered to the individual. Alternatively, an agent which increases the activity of the protein of SEQ ID NO:316 may be administered to the
35 individual. Such agents may be identified by contacting the protein of SEQ ID NO:316 or a cell or preparation containing the protein of SEQ ID NO:316 with a test agent and assaying whether the

test agent increases the activity of the protein. For example, the test agent may be a chemical compound or a polypeptide or peptide.

Alternatively, the activity of the protein of SEQ ID NO:316 may be decreased by administering an agent which interferes with such activity to an individual. Agents which interfere with the activity of the protein of SEQ ID NO:316 may be identified by contacting the protein of SEQ ID NO:316 or a cell or preparation containing the protein of SEQ ID NO:316 with a test agent and assaying whether the test agent decreases the activity of the protein. For example, the agent may be a chemical compound, a polypeptide or peptide, an antibody, or a nucleic acid such as an antisense nucleic acid or a triple helix-forming nucleic acid.

10 In one embodiment, the invention relates to methods and compositions using the protein of the invention or part thereof as a marker protein to selectively identify the source of a sample as, for example, brain, or to distinguish between two or more possible sources of a sample on the basis of the level of the protein of SEQ ID NO:316 in the sample. For example, the protein of SEQ ID NO:316 or fragments thereof may be used to generate antibodies using any techniques known to those skilled in the art, including those described therein. Such antibodies may then be used to identify tissues of unknown origin, for example, forensic samples, differentiated tumor tissue that has metastasized to foreign bodily sites, or to differentiate different tissue types in a tissue cross-section using immunochemistry. In such methods a sample is contacted with the antibody, which may be detectably labeled, under conditions which facilitate antibody binding. The level of antibody binding to the test sample is measured and compared to the level of binding to control cells from brain or tissues other than brain to determine whether the test sample is from brain. Alternatively, the level of the protein of SEQ ID NO:316 in a test sample may be measured by determining the level of RNA encoding the protein of SEQ ID NO:316 in the test sample. RNA levels may be measured using nucleic acid arrays or using techniques such as in situ hybridization, Northern blots, dot blots or other techniques familiar to those skilled in the art. If desired, an amplification reaction, such as a PCR reaction, may be performed on the nucleic acid sample prior to analysis. The level of RNA in the test sample is compared to RNA levels in control cells from brain or tissues other than brain to determine whether the test sample is from brain.

In another embodiment, antibodies to the protein of the invention or part thereof may be used for detection, enrichment, or purification of cells expressing the protein of SEQ ID NO:316, including using methods known to those skilled in the art. For example, an antibody against the protein of SEQ ID NO:316 or a fragment thereof may be fixed to a solid support, such as a chromatography matrix. A preparation containing cells expressing the protein of SEQ ID NO:316 is placed in contact with the antibody under conditions which facilitate binding to the antibody. The support is washed and then the cells are released from the support by contacting the support with agents which cause the cells to dissociate from the antibody.

In another embodiment of the present invention, the protein of SEQ ID NO:316 or a fragment thereof may be used to diagnose disorders associated with altered expression of the protein of SEQ ID NO:316. In such techniques, the level of the protein of SEQ ID NO:316 in an ill individual is measured using techniques such as those described herein. The level of the protein of
5 SEQ ID NO:316 in the ill individual is compared to the level in normal individuals to determine whether the individual has a level of the protein of SEQ ID NO:316 which is associated with disease.

Protein of SEQ ID NO:255 (106-037-1-0-E9-CS.cor)

The protein of SEQ ID NO:255, encoded by the cDNA of SEQ ID NO:14, is strongly
10 expressed in the liver and testis and shows extensive homology to human lactate dehydrogenase-A protein (LDH-A or M chain) (Chung F.Z. et al., Biochem. J. 231:537-541(1985); SwissProt accession number P00338). The protein of SEQ ID NO:255 is also homologous to lactate dehydrogenase A from many vertebrates. The 381-amino-acid-long protein of SEQ ID NO:255 displays a Prosite motif corresponding to lactate dehydrogenase from positions 71 to 380. In
15 addition, the active site LGEHGDS, where H is the active site residue, is present in the protein of the invention (positions 239 to 245). The protein of the invention also contains an additional 50 N-terminal amino acids not found in other lactate dehydrogenase A proteins. This N-terminal extension contains a signal peptide (cleavage site at position 34 of the protein of invention) that may allow the export of the protein to the extracellular domain or define a particular subcellular
20 localization. Alternatively, the initiation start codon could be at position 26 or 50 of the protein of SEQ ID NO:255.

Lactate dehydrogenase (LDH) is an enzyme which dehydrogenates lactic acid into pyruvic acid in conjunction with the hydrogen acceptor NAD⁺, and which exists in a wide variety of animal tissues and microorganisms as an enzyme serving to produce lactic acid
25 from pyruvic acid in the glycolytic pathway (Abad-Zapatero C. et al. J. Mol. Biol. 198:445-467(1987)). It is known that in vertebrates there are three isozymes of LDH: the M form (LDH-A), found predominantly in muscle tissues; the H form (LDH-B), found in heart muscle, and the X form (LDH-C), found only in the spermatozoa of mammals and birds. In birds and crocodilian eye lenses, LDH-B serves as a structural protein and is known as
30 epsilon-crystallin (Hendriks W. et al. Proc. Natl. Acad. Sci. U.S.A. 85:7114-7118(1988)).

LDH has been used extensively in the field of clinical test reagents for a number of purposes. For example, it has been used as a coupling enzyme to determine the enzymatic activity of various amino-transferases, such as alanine aminotransferase (ALT), which is ultimately detected by UV spectrometry of the produced pyruvic acid. This use of LDH
35 has been widely adopted as a clinical test, because amino-transferases are enzymes which

show high activity in liver, heart, kidney, etc. and show remarkable increases in serum in association with various diseases. LDH has also been used as a coupling enzyme to help determine the level of substrates such as urea, as the enzyme promotes the conversion of such substances into pyruvic acid which can be detected by UV spectrometry.

5 Lactate dehydrogenase is also a widely used marker for heart disease and other conditions. For example, levels of LD-1 are elevated in the presence of myocardial infarction and in other conditions such as leukemia. Levels of lactate dehydrogenase start to increase 24 to 48 hours after occlusion of the coronary artery, peak in 3 to 6 days, and return to normal in 8 to 14 days. In addition, levels of LD-1 are elevated 10 to 12 hours
10 after the acute myocardial infarction, peak in 2 to 3 days, and return to normal in approximately 7 to 10 days. Thus, measurement of the level of lactate dehydrogenase allows a prolonged retrospective diagnosis of myocardial infarction. Further, while the amount of LD-2 in the blood is usually higher than the amount of LD-1, patients with acute myocardial infarction have more LD-1 than LD-2. This "flipped ratio" usually returns to
15 normal in 7 to 10 days. An elevated level of LD-1 with a flipped ratio has a sensitivity and specificity of approximately 75% to 90% for detection of acute myocardial infarction.

Elevated LDH levels have also been used as a prognostic indicator for cancers such as small cell lung carcinoma. Specifically, elevated levels of LDH indicate a poor prognosis for such diseases (Kawahara, et al., (1997) Jpn J Clin Oncol. 1997 Jun;27(3):158-
20 65).

LDH expression in cells has also been shown to be induced by interleukin-1 alpha, a major cytokine associated with, e.g., inflammation (Nehar et al. (1998) Biol Reprod Dec;59(6):1425-32).

Islet beta-cells express low levels of lactate dehydrogenase and have high glycerol
25 phosphate dehydrogenase activity. The effects on glucose metabolism and insulin secretion of acute overexpression of the skeletal muscle isoform of lactate dehydrogenase (LDH)-A in these cells have been studied by Ainscow EK et al. (Diabetes 2000 Jul;49(7):1149). The results of these studies have shown that overexpression of LDH activity interferes with normal glucose metabolism and insulin secretion in islet beta cells, and it may therefore be
30 directly responsible for insulin secretory defects in some forms of type 2 diabetes. These results also reinforce the view that glucose-derived pyruvate metabolism in the mitochondria is critical for glucose-stimulated insulin secretion in beta cells. Other data show that an overexpression of lactate dehydrogenase A attenuates glucose-induced insulin secretion in stable MIN-6 beta-cell lines, which normally express low levels of L-lactate

dehydrogenase (Zhao C, Rutter GA FEBS Lett. 1998 Jul 3;430(3):213-6). Low LDH activity thus appears to be important in beta-cell glucose sensing.

Analysis of the LDH isoenzyme pattern in CSF fluid has also been shown to be helpful in the evaluation of CNS involvement in patients with hematologic malignancies (Lossos IS, et al.

5 Cancer. 2000 Apr 1;88(7):1599-604).

It is believed that the protein of SEQ ID NO:255 is a lactate dehydrogenase protein, most likely of the LDH-A or M subtype. The activity of the present protein can be assessed using any standard method for detecting lactate dehydrogenase enzyme activity, including those involving the UV detection of pyruvate, a product of LDH-catalyzed enzymatic reactions.

10 In one embodiment, the polypeptides and polynucleotides of the invention are used to detect testis and liver tissue, as well as cells derived from these tissues. For example, nucleic acids and proteins of the invention can be labeled isotopically or chemically, using methods known to those skilled in the art, and used as probes in northern blots, far-western blots and *in situ* hybridization experiments. An ability to detect specific cell types is useful, e.g. for the determination of the
15 history of tumor cells, as well as for the identification of cells and tissues for histological studies.

In another embodiment, the present protein can be used in any of a variety of clinical assays involving LDH enzymes. For example, the protein can be used as a coupling enzyme to determine the enzymatic activity of various amino-transferases, such as alanine aminotransferase (ALT), as detected by UV spectrometry of the produced pyruvic acid. Such assays have significant clinical
20 utility, as amino-transferases are enzymes which show high activity in liver, heart, kidney, etc. and show remarkable increases in serum in association with various diseases. The protein of the invention can also be used as a coupling enzyme to help determine the level of substrates such as urea, as the enzyme promotes the conversion of such substances into pyruvic acid which can be detected by UV spectrometry.

25 In another embodiment, the present protein can be used to identify ingredients for cosmetic formulations. Specifically, enhancers of lactate dehydrogenase can be included in cosmetic compositions to stimulate keratinocyte proliferation and collagen synthesis in cutaneous tissues. The inhibitors can be combined with other active ingredients such as pyruvic acid, acetic acid, acetoacetic acid, beta-hydroxybutyric acid, Krebs cycle pathway metabolites, aliphatic saturated or
30 unsaturated fatty acids containing from 8 to 26 carbon atoms, omega-hydroxy acids containing from 22 to 34 carbon atoms, glutamic acid, glutamine, valine, alanine, leucine, and mixtures thereof (see, e.g., US Patent 5,853,742, the disclosure of which is hereby incorporated by reference in its entirety).

In another embodiment, the present invention provides methods for treating or preventing
35 cancer, e.g., by inhibiting lactate dehydrogenase activity in cells, preferably specifically the cancer cells, of a patient. The expression or activity of lactate dehydrogenase can be inhibited using any of a large number of agents, including, but not limited to, antibodies, antisense molecules, ribozymes,

and heterologous molecules that inhibit the expression or activity of the lactate dehydrogenase in the cancer cells of the patient. In one embodiment, lactate dehydrogenase that has been obtained from a primate, or anti-lactate dehydrogenase antibodies obtained from a mammal as a result of the parenteral administration of primate lactate dehydrogenase to said mammal, is parenterally

5 administered to human cancer patients. Antibodies derived from the protein of the invention or part thereof can also be used to inhibit cancer cell development as described in US Patent No. 4,620,972.

Analysis of the LDH isoenzyme pattern in CSF fluid has been shown to be helpful in the evaluation of CNS involvement in patients with hematologic malignancies (Lossos IS, et al. Cancer. 2000 Apr 1;88(7):1599-604). Thus, in another embodiment, the protein of SEQ ID NO:255 can be
10 used to develop assays to monitor the LDH isoenzyme activity in CSF fluid, thereby improving the sensitivity of CSF cytology. This assay may be derived, e.g., from the methods described by Short S. et al. (J Biol Chem. 2000 Apr 28;275(17):12963-9).

In another embodiment, the protein of SEQ ID NO:255 is used to detect and/or treat insulin secretory defects in some forms of type 2 diabetes. For example, various evidence indicates that
15 LDH overexpression may be involved in certain types of diabetes. Therefore, the detection of an elevated level of LDH in a patient, e.g. in pancreatic islet cells of a patient, can be used as an indication that the patient has diabetes, or is at risk of developing diabetes. Similarly, methods of inhibiting the expression or activity of LDH in those cells, e.g. using antibodies, antisense sequences, or heterologous compounds that inhibit the expression or activity of LDH, can be used to
20 treat or prevent diabetes.

In another embodiment, the protein of the invention can be used to eliminate endogenous pyruvic acid in cells in vitro or in vivo.

In another embodiment, the expression of the present protein is used as a marker for interleukin 1, e.g. IL-1 alpha, activity in cells or in a patient. Specifically, as it has been shown that
25 LDH expression is induced by IL-1 alpha, then the expression, or elevated expression, of the present protein can be used as a marker for the action of IL-1 on the cell. As IL-1 has been implicated in a number of physiological processes, including inflammation and more specifically in deleterious processes such as arthritis and autoimmune disorders, the present protein can serve as a marker for the presence of such disorders, or for a predisposition for the disorders.

30 In another embodiment, the present protein is used to detect heart disease and other diseases in patients. For example, levels of LDH are known to rise following myocardial infarction and other heart ailments. Accordingly, the detection of an elevated level of the protein of the invention, alone or in view of the levels of other proteins such as other LDH isozymes, can be used as an indicator of a heart attack or other diseases, including
35 leukemia. The levels of LDH can be assessed in any tissue or biological sample, including, but not limited to, serum, and can be detecting using any standard method, including, but not limited to, immunoassays and assays for LDH enzyme activity.

In another embodiment, the present protein is used to determine a prognosis for any of a number of diseases, including cancers such as small cell lung carcinoma. For example, the level of the present protein is detected in the serum of a patient suffering from cancer, wherein the detection of a decreased level of expression or activity of the protein indicates a worse prognosis for the patient compared to the prognosis in a patient with a normal level of the protein activity or expression.

Proteins of SEQ ID NOs:243, 253 (internal designation numbers 105-016-1-0-D3-CS and 105-095-2-0-G11-CS)

The 331-amino-acid- long protein of SEQ ID NO:243, encoded by the cDNA of SEQ ID NO:2, is found in prostate and in fetal brain and is homologous to a secreted human protein (Genseq accession number Y59685). In addition, this protein is highly homologous to the putative glycerophosphodiester phosphodiesterase (GP-PDE) MIR16 (Membrane Interacting protein of RGS16) protein (SPTREMBLNEW SPTREMBL SWISSPROT accession number AAF65234) encoded by the cDNA of GENPEPT GENPEPTNEW accession number AF212862; in fact, the protein of the invention is a likely variant of the MIR16 protein. Furthermore, a BLAST search with the amino acid sequence of SEQ ID NO:243 indicates that the protein of the invention is homologous to GP-PDEs of *E.coli* (SWISSPROT accession numbers P09394 and P10908) and *Haemophilus influenzae* (SWISSPROT accession number Q06282). The protein of SEQ ID NO:243 displays 2 candidate membrane-spanning segments, from amino acids 7 to 27 and 258 to 278, and a putative signal peptide from amino acids 19 to 24. Finally, the protein of the invention has two putative *N*-glycosylation sites: asparagine residues at positions 168 and 198 (Zheng *et al.*, Proc. Natl. Acad. Sci. 97 :3999-4004 (2000)).

The cDNA of SEQ ID NO:2 differs from the cDNA of GENPEPT GENPEPTNEW accession number AF212862 by its extended 5' and 3' termini, and from the cDNA of SEQ ID NO:12 by polymorphisms and alternate splicings.

The MIR16 (Membrane Interacting protein of RGS16) protein, which is homologous to the protein of the invention, was identified in a yeast two-hybrid screen of a pituitary cell cDNA library using the RGS16 (Regulator of G protein Signaling) protein as bait (Zheng *et al.*, Proc. Natl. Acad. Sci. 97:3999-4004 (1999)). and Sasaki, J. Bacteriol. 175:4569-4571 (1993); Zheng *et al.*, *ibid.*). Remarkably, the GP-PDE from *Haemophilus influenzae* (also called protein D) which is 67% identical to the periplasmic GP-PDE of *E.coli*, presents affinity for human immunoglobulin D (Janson *et al.*, Infect. Immun. 62:4848-854 (1994)).

From sequence alignments, it can be seen that the N-terminal region of MIR16 (amino acids 70-150), immediately after the putative signal peptide, is highly conserved (40-61% similarity), suggesting that it may contain residues critical for catalytic activity, i.e., the catalytic site. GP-PDEs hydrolyze deacetylated phospholipid GPs, such as glycerophosphocholine (GPC)

and glycerophosphoethanolamine, to sn-glycerol-3-phosphate (G3P) and the corresponding alcohols (Zheng *et al.*, *ibid.*). The putative enzymatic activity of MIR16 and its interaction with RGS16 suggest that it may play important roles in lipid metabolism and in G protein signaling. As shown in northern blot experiments, the MIR16 mRNA is highly transcribed in heart, liver, kidney, testis and brain. The observed expression of MIR16 in the brain is consistent with the above-described expression of the protein of the invention in the fetal brain.

It is believed that the proteins of SEQ ID NOs:243 and 253 or part thereof are members of the glycerophosphodiester phosphodiesterase protein family, interact with the RGS16 protein and, as such, play important roles in both lipid metabolism and in G protein signaling. Preferred polypeptides of the invention are polypeptides comprising the amino acids of SEQ ID NO:243 from positions 7 to 27, 19 to 24 and 258 to 278. Other preferred polypeptides of the invention are fragments of SEQ ID NO:243 or 253 having any of the biological activities described herein. Additional preferred polypeptides are those that comprise asparagine residues at positions 168 and/or 198.

The invention first relates to methods and compositions using cDNAs of SEQ ID NO:2 or 12 or part thereof, and proteins of the invention SEQ ID NO:243 or 253 or part thereof to identify specific cell types, preferably from prostate or fetal brain. For example, nucleic acids and proteins of the invention are labeled isotopically or chemically following methods known to those skilled in the art, and further used as probes in northern blots, far-western blots and *in situ* hybridization detection experiments. An ability to detect specific cell types is useful, e.g. for the determination of the history of tumor cells, as well as for the identification of cells and tissues for histological studies.

Any of a number of *in vitro* assays can be used to detect SEQ ID NO:243 or 253 protein activity, for example for *in vitro* screening of modulators of protein activity. Preferably cDNA encoding the protein of the invention is cloned in a prokaryotic expression vector, according to methods known to those skilled in the art. Briefly, the GP-PDE activity of the recombinant protein is analyzed by a coupled spectrophotometric assay as described by Larson and collaborators and adapted by Cameron and collaborators (Larson *et al.*, *J. Biol. Chem.* 258 :5426-5432 (1983); Cameron *et al.*, *Infect. Immun.* 66 :5763-5770 (1998)). Such enzymatic activity may be measured *in vitro* in the presence of modulating drugs.

Another embodiment of the present invention relates to methods of using the protein of the invention or part thereof to purify or specifically bind to human immunoglobulin D. Several immunoglobulin (Ig) binding bacterial cell wall proteins have been isolated and/or cloned during the last two decades. The best characterized of these are protein A of *Staphylococcus aureus* (which binds to human IgG subclasses 1, 2 and 4, IgG of several mammals species, and in some instances human Ig of classes A, M, E), and protein G of group G beta-hemolytic streptococci (which binds to all human IgG subclasses and which also displays a wider binding spectrum for

animal IgG than protein A). IgD binds to neither protein A nor protein G. Consequently, it is of great interest to identify new proteins capable of binding IgD, thereby allowing its separation and purification. In addition, IgD binding proteins can also be used in immunoprecipitation procedures with IgD, as are routinely performed with proteins A and G in the case of IgG. The binding and
5 purification of IgD using the protein of the invention can be accomplished in any of a number of ways, for example by generating a fusion protein or polypeptide in which the protein of the invention or part thereof, is combined with another protein by the use of a recombinant DNA molecule. The resulting fusion product including the protein of the invention or part thereof is then covalently, or by any other means, bound to a protein, carbohydrate or matrix (such as gold,
10 "Sephadex" particles, polymeric surfaces). Such a complex is very useful for IgDs immobilization and consecutive immunoprecipitations in batch. Similar assays for binding of protein D (GP-PDE) of *Haemophilus influenzae* and IgD are described in the US Patent No. 6,025,484.

Another embodiment of the invention relates to compositions and methods using the protein of the invention, or part thereof, as GP-PDE enzymes to hydrolyze deacylated phospholipids (GPs),
15 such as glycerophosphocholine (GPC) and glycerophosphoethanolamine, to sn-glycerol-3-phosphate (G3P) and the corresponding alcohols. First, this enzymatic activity, which belongs to the class of specific phospholipase D, makes the protein of the invention very useful to study biological membranes and their phospholipidic components. Moreover, as glycerophospholipids are major components of the lipidic bilayer, elimination of their hydrophilic moiety using the GP-
20 PDE activity of the protein of the invention would likely modify the structure and consequently the permeability of eukaryotic cell membranes. Such modifications could improve the transfection efficiency of eukaryotic cells, in vitro or in vivo. Typically, in such embodiments the purified protein of SEQ ID NOs:243 or 253 is administrated to cells; purified proteins of the invention can be obtained in any of a number ways, for example by inserting the cDNA encoding the proteins into
25 a prokaryotic expression vector using any technique known to those skilled in the art. The recombinant protein produced and purified in the prokaryotic system is then added to an *in vitro* culture of eukaryote cells before or during transfection. The recombinant protein of the invention can also be used to increase the efficiency of cell transfection *in vivo*, most notably in the case of gene therapy. For example, tumoral masses are very often resistant to transfection, and the protein
30 of the invention would likely provide an effective way to facilitate the introduction of cytotoxic genes (such as pro-apoptotic genes) or antitumoral drugs in solid tumors.

Still another embodiment of the protein of the invention relates to methods and compositions to diagnose, treat, and prevent disorders associated with excess glutamate signaling in the brain. As described above, the MIR16 protein interacts physically with the RGS16 protein
35 (Regulator of G protein Signaling 16). Receptors of many hormones use heterotrimeric G proteins for signal transduction after ligand binding (for a review, see Neer, Cell 80 :249-257 (1995)). Among these receptors are metabotropic glutamate receptors (mGluRs). These receptors, which are

expressed in the brain, like the protein of the invention, are a novel family of cloned G-protein-coupled receptors (Schoepp and Conn, Trends Pharmacol. Sci. 14:13-20 (1993)). Endogenous glutamate, by activating the mGluR1 receptor (and also NMDA and AMPA receptors), may contribute to the brain damage occurring acutely after epilepsy, cerebral ischemia or traumatic brain injury. It may also contribute to chronic neurodegeneration in such disorders as amyotrophic lateral sclerosis and Huntington's chorea (Meldrum, J. Nutr. 130(4S Suppl):1007S-1015S (2000)).

The invention thus relates to methods and compositions using cDNAs of SEQ ID NO:2 or 12 or part thereof, and proteins of SEQ ID NO:243 or 253 or part thereof, to diagnose, treat, or prevent disorders associated with excess glutamate signaling in the brain. Specifically, the level of activity or expression of the proteins can be correlated with the level of glutamate signaling, or with the glutamate-signaling associated brain damage involved in epilepsy, cerebral ischemia, traumatic brain damage, ALS, or Huntington's chorea, or with any other G-protein associated physiological process or disease or condition. For situations where the level of the expression or activity of the protein is positively correlated with such signaling or with the presence of a disease or condition, the signaling, disease or condition can be detected using any of a number of tools for detecting protein expression or activity, including northern blots, far-western blots and *in situ* hybridization experiments, where an elevated level of the protein, protein activity, or nucleic acid of the invention indicates the presence of the disease, condition, or signaling process. Further, such diseases or conditions can be treated or prevented, or such signaling pathways can be inhibited, using compounds that inhibit the expression or activity of the protein, such as antibodies, antisense molecules, ribozymes, dominant negative forms of the protein, or any heterologous molecule that inhibits protein activity or expression. Alternatively, where the expression or activity of the protein of the invention is negatively associated with the signaling pathway, disease or condition, a detection of a decreased level of expression or activity of the protein can be used to indicate the presence of the disease, condition, or pathway. Further, in such cases, the disease or condition can be treated or prevented, or the pathway be inhibited, using any compound that increases the activity or level of the protein, such as nucleic acids encoding the protein, the protein itself, or heterologous compounds that cause an increase in the level of protein expression or activity.

Protein of SEQ ID NO:386 (internal designation 105-037-4-O-H12-CS)

The protein of SEQ ID NO:386, encoded by the cDNA of SEQ ID NO:145, is strongly expressed in the fetal brain and uterus. The 207-amino-acid-long protein of SEQ ID NO:386 displays pfam SPRY domains from positions 85 to 205.

SPRY domains have been found in a number of proteins involved in multiple cellular and developmental processes. For example, the Midline-1/FXY family of proteins has been shown to associate with microtubules, and has been implicated in human diseases, such as Opitz Syndrome, a congenital disorder characterized by multiple developmental abnormalities (see, e.g., Cainarca, et

al., (1999) Hum Mol Genet 8(8):1387-96). In addition, the cytoplasmic Marenostirin/Pyrin protein has been demonstrated to be the cause of Familial Mediterranean fever, an autosomal recessive disorder characterized by fever and serositis (Nat Genet 1997 Sep;17(1):25-31). Other SPRY proteins include SplA, a serine protease from Staphylococcus aureus, and butyrophilin, a major milk protein. Another family of proteins known to contain the SPRY domain are the Ryanodine receptors (RyRs).

Ryanodine receptors play an important role in Ca^{2+} signaling in muscle and non muscle cells by releasing Ca^{2+} from intracellular stores. For example, these receptors are centrally important in excitation-contraction (e-c) coupling, which occurs at specialized regions where the sarcoplasmic reticulum (SR), containing the ryanodin receptors, and the plasma membrane/transverse-tubule system form junctions. RyRs are also thought to play some role in maintaining the structural integrity of the SR-T-tubule junctions. RyR is apparently unable to carry out the requisite functions associated with e-c coupling by itself, however, because it forms interactions with other macromolecules at the triad junction. For example, two small proteins, calmodulin and FKBP12, are believed to modulate RyR at the triad junction.

It is believed that mammalian tissues express three different RyR isoforms, comprising four 560-kDa (RyR polypeptide) and four 12-kDa (FK506 binding protein) subunits. It is believed that these large protein complexes conduct monovalent and divalent cations and are capable of multiple interactions with other molecules. The subunits of the protein complexes include small diffusible endogenous effector molecules including Ca^{2+} , Mg^{2+} , adenine nucleotides, sufhdyryl modifying reagents (glutathione, NO, and NO adducts) and lipid intermediates, and proteins such as protein kinases and phosphatases, calmodulin, immunophilins (FK506 binding proteins), and in skeletal muscle the dihydropyridine receptor. The RyR from skeletal muscle is the major calcium release channel for that tissue, and the most intensively studied of the three genetic isoforms detected thus far in mammalian species. The other two RyR isoforms are often referred to as the 'heart' and 'brain' forms, but the actual cell and tissue distribution of the isoforms is complex.

Because of their multiple ligand interactions, ryanodin receptors constitute an important, potentially rich pharmacological target for controlling cellular functions. Ca^{2+} release channel activity is modulated by many endogenous effectors, including Ca^{2+} , ATP, Mg^{2+} , and calmodulin. In addition, many exogenous effectors, including caffeine, local anesthetics, and polyamines, also modify channel activity. For example, tetracaine, procaine, benzocaine, and lidocaine inhibit Ca^{2+} release from the SR. They appear to interact with a specific site(s) located on the RYR, affecting both ryanodin-binding and single channel activities (Shoshan-Barmatz et al. 1993; J. Membr. Biol.; 133; 171-181).

The importance of intracellular calcium as a second messenger in cellular signal transduction processes is well established. Alterations in intracellular Ca^{2+} homeostasis have profound effects on many cell functions, including secretion, contraction-relaxation, motility,

metabolism, protein synthesis, modification and folding, gene expression, cell-cycle progression and apoptosis. A major source of cytoplasmic calcium is from intracellular storehouses located in the endoplasmic reticulum, or in muscle, within the sarcoplasmic reticulum (SR).

Given that cellular Ca^{2+} handling is an important factor in the control of neuronal metabolism and electrical activity, abnormalities of intracellular Ca^{2+} channels might be expected to contribute to some forms of epilepsy or to anoxic brain damage following an episode of cerebral ischemia. Cell loss is said to be a characteristic feature of degenerative brain disorders, including Alzheimer's disease. It is well established that neuronal cell death may be secondary to an abnormal elevation of cytoplasmic Ca^{2+} , particularly that associated with activation of excitatory glutamate receptors (e.g., in epilepsy). This strongly suggests that the release of stored Ca^{2+} contributes to nerve cell damage and cell death in various circumstances.

It is believed that the protein of SEQ ID NO:386 is functionally related to other SPRY-containing proteins, such as the ryanodine receptors, Marenostin/Pyrin, SplA, Midline-1/FXY, and butyrophilin. Accordingly, it is thus believed that the present protein is associated with the release of Ca^{2+} from intracellular Ca^{2+} -storing organelles, like the endoplasmic reticulum and, in muscle, the sarcoplasmic reticulum (SR), as well as being involved in microtubule binding. Preferred polypeptides of the invention are any fragments of SEQ ID NO:386 having any of the biological activities described herein.

In one embodiment, the present protein and nucleic acids can be used to specifically detect cells of the fetal brain and uterus, as the protein is overexpressed in these tissues. For example, the protein of the invention or part thereof may be used to synthesize specific antibodies using any technique known to those skilled in the art. Such tissue-specific antibodies may then be used to identify tissues of unknown origin, such as in forensic samples, differentiated tumor tissue that has metastasized to foreign bodily sites, etc., or to differentiate different tissue types in a tissue cross-section using immunochemistry. The protein can also be used to specifically label microtubules in cells.

In another embodiment, the protein of the invention or part thereof may be used in regulating intracellular Ca^{2+} levels. As alterations in intracellular Ca^{2+} homeostasis have profound effects on many cell functions, including secretion, contraction-relaxation, motility, metabolism, protein synthesis, modification and folding, gene expression, cell-cycle progression and apoptosis, the ability to modulate intracellular Ca^{2+} levels provides a tool to alter any of these cellular functions, in vitro or in vivo. Such an ability has wide utility for a large number of applications, for example to manipulate the behavior (e.g. growth rate, secretion, survival, etc.) of cells grown in vitro, as well as to treat, prevent, or diagnose any of a number of diseases associated with altered Ca^{2+} signaling in vivo. The activity or expression of the protein of the invention can be modulated in any of a large number of ways, for example by administering to cells or to a patient the protein itself, a polynucleotide encoding the protein, antibodies, antisense sequences, dominant negative

forms of the protein, compounds that alter the expression or activity of the protein, etc. The effect of any such agent on calcium flux in cells can be detected using standard methods, including by studying the permeation of Ca^{2+} release through endoplasmic reticulum (ER) and sarcoplasmic reticulum (SR) channels using tracers, light scattering and fluorescence quenching, and channel
5 reconstitution in planar bilayer. In addition, targeted recombinant photoproteins can provide direct measurements of organellar Ca^{2+} (Montero et al.; 1995; EMBO J.; 14, 5467-5475).

The invention further relates to methods and compositions using the protein of the invention or part thereof to diagnose, prevent and/or treat several disorders in which the activity or recognition of ryanodin receptors, is impaired or excessive. These disorders include, but are not
10 limited to, neurodegenerative diseases, cardiovascular disorders, severe myasthenia, malignant hyperthermia, epilepsy, and central core disease. For example, in patients with severe myasthenia, the level of anti-RyR antibodies has been directly related to the severity of the disease (Skeie et al., 1996: Eur. J. Neurol. 3; 136-140). There is also some evidence to suggest that RyR abnormalities are a primary cause of many types of cardiac disease. In addition, the protein of the invention can
15 be used to diagnose other diseases associated with SPRY-protein dysfunction, such as Familial Mediterranean fever and Opitz syndrome. Finally, as SPRY containing proteins have been implicated in embryonic development (e.g. the Midline1 protein), the protein and nucleic acids of the invention can be used to detect developmental disorders, as the detection of a mutation in the gene encoding SEQ ID NO:386, or a detection of abnormal gene expression in a fetus, can be used
20 to indicate the presence of a developmental abnormality. For example, as the protein of SEQ ID NO:386 is strongly expressed in the fetal brain, it is likely that the protein plays a role in the normal development of the brain in utero.

The present invention also relates to diagnostic assays for detecting altered levels of the protein of SEQ ID NO:386 in various tissues, as over-expression of the protein compared to normal
25 control tissue samples can indicate the presence of certain disease conditions such as neurodegenerative disorders, cardiovascular disorders, severe myasthenia, malignant hyperthermia, epilepsy, and central core disease. Assays used to detect levels of the polypeptide of the present invention in a sample derived from a host are well-known to those of skill in the art and include radioimmunoassays competitive-binding assays, Western Blot analysis and ELISA assays.

30 Proteins SEQ ID NOs:283 and 286 (internal designations 174-38-1-0B6-CS LA and 174-41-1-0-A6-CS LA)

The protein of SEQ ID NO:283, encoded by the cDNA of SEQ ID NO:42, is overexpressed in salivary glands and to a lesser extent in bone marrow, and shows homology over the C-terminal length to the immunoglobulin (Ig) protein superfamily, which is conserved among eukaryotes
35 (including rabbit, rodents and human). In particular, the 468-amino-acid-long protein of the invention, which is similar in size to the constant chain of Ig related proteins, displays two pfam

conserved immunoglobulin domains, from position 205 to 285 and from position 318 to 384, which are known to be involved in the basic structure of the light and heavy constant chains of immunoglobulins. It is known (Orr H.T., Nature 282:266-270(1979)) that the Ig constant chain domains and a single extracellular domain in each type of MHC chain are closely related, sharing
5 over one hundred amino-acids of homology. All members of the Ig related superfamily, including the MHC class I alpha chain and beta-2-microglobulin, as well as the MHC class II alpha and beta chains, display the prosite conserved characteristic pattern around the C-terminal cysteine ([FY]-x-C-x-[VA]-x-H). This cysteine is involved in the disulfide bond between the light and heavy chains, and is also found in the protein of the invention (position 380 to 386). The protein of the invention
10 also exhibits an emotif Ig and Major Histocompatibility Complex protein signature from positions 319 to 336. In addition, the protein of the invention displays homology with tapasin (GeneBank No. AF009510), a chaperone-like protein closely associated with TAP-binding proteins, which is well conserved among eukaryotes (chicken, rodents and human). Tapasin has been shown to increase the efficiency of antigen processing and presentation by mediating the association of MHC
15 complex proteins with TAP proteins to the endoplasmic reticulum and to the cell surface during immune response (for review see Abele, R. and Tampé, R., Bioch. et Biophysica Acta, 1999). In addition, the protein of the invention displays two transmembrane domains from positions 199 to 219 and from positions 406 to 426, a hydrophobic profile similar in amino acid position to the hydrophobic stretch of amino acids of human and mouse tapasin (Suling L., J. Biol. Chem.,
20 274:8649-8654, 1999), and a secreted signal peptide from position 9 to 23. Both signatures are largely present in Ig related proteins such as secreted antibodies or antigen presenting proteins. The invention also encompasses a variant (SEQ ID NO:286) of SEQ ID NO:283, encoded by the cDNA of SEQ ID NO:45. The protein of SEQ ID No:286 is a 442-amino-acid-long protein with a C-terminal shorter end of 26 amino-acids compared to the protein of SEQ ID NO:283. The variant of
25 SEQ ID NO:286, which results from a frameshift (position 1445 in SEQ ID NO:45) in the coding sequence that leads to a stop codon in the corresponding protein, displays characteristics identical to those described above in terms of motifs, Ig signatures, function, and potential uses.

The immunoglobulin (Ig) gene superfamily comprises a large number of cell surface glycoproteins that share sequence homology with the V and C domains of antibody heavy and light
30 chains. These molecules function as receptors for antigens, immunoglobulins and cytokines as well as adhesion molecules, and play important roles in regulating the complex cell interactions that occur within the immune system (A. F. Williams et al., Annu. Rev. Immuno. 6:381-405, 1988, T. Hunkapiller et al., Adv. Immunol. 44:1-63, 1989; for a short review see also Prosite entry PS00290)

The introduction of an antigen into a host initiates a series of events culminating in an
35 immune response. In addition, self-antigens can result in immunological tolerance or activation of an immune response against self-antigens. A major portion of the immune response is regulated by presentation of antigen by major histocompatibility complex molecules. MHC molecules bind to

peptide fragments derived from antigens to form complexes that are recognized by T cell receptors on the surface of T cells, giving rise to the phenomenon of MHC-restricted T cell recognition. The ability of a host to react to a given antigen (responsiveness) is influenced by the spectrum of MHC molecules expressed by the host. Responsiveness correlates with the ability of specific peptide fragments to bind to particular MHC molecules.

There are two types of MHC molecules, class I and class II, each of which comprise two chains. In class I [2], the alpha chain is composed of three extracellular domains, a transmembrane region, and a cytoplasmic tail. The beta chain (beta-2-microglobulin) is composed of a single extracellular domain. In class II [3], both the alpha and the beta chains are composed of two extracellular domains, a transmembrane region and a cytoplasmic tail. MHC class I molecules are expressed on the surface of all cells, and MHC class II molecules are expressed on the surface of antigen presenting cells. MHC class II molecules bind to peptides derived from proteins made outside of an antigen presenting cell. In contrast, MHC class I molecules bind to peptides derived from proteins made inside a cell. In order to present peptide in the context of a class II molecule, an antigen presenting cell phagocytoses an antigen into an intracellular vesicle, in which the antigen is cleaved, bound to an MHC class II molecule, and then returned to the surface of the antigen presenting cell.

Major histocompatibility complex (MHC) class I molecules present antigenic peptides to CD8 T cells (Townsend, A. et al., Nature:340,443-448)). The peptides are generated in the cytosol and then translocated across the membrane of the endoplasmic reticulum by the transporter associated with antigen processing (TAP). TAP is a trimeric complex consisting of TAP1, TAP2, and tapasin (TAP-A). TAP1 and TAP2 are required for the peptide transport. Tapasin mediates the interaction of MHC class I HC-beta-2 microglobulin with TAP, and this interaction is essential for peptide loading onto MHC class I HC-beta-2-microglobulin (Suling et al., J. Biol. Chem., 274:8649-8654). T cell receptors (TCRs) are the second antigen recognition molecules, and recognize antigens that are bound by MHC molecules. Recognition of MHC complexed with peptide (MHC-peptide complex) by TCR can effect the activity of the T cell bearing the TCR. Thus, MHC-peptide complexes are important in the regulation of T cell activity and, thus, in regulating an immune response.

Human cytomegalovirus (HCMV) is a betaherpesvirus which causes clinically serious disease in immunocompromised and immunosuppressed adults, as well as in some infants infected in utero or perinatally (Alford, C. A., and W. J. Britt. 1990. Cytomegalovirus, p. 1981-2010. In D. M. Knipe and B. N. Fields (ed.), Virology, 2nd ed. Raven press, New York). In human cytomegalovirus (HCMV)-infected cells, expression of the cellular major histocompatibility complex (MHC) class I heavy chains is down-regulated, where down-regulation is defined as reduction in either synthesis, stability or surface expression of MHC class I heavy chains. A similar phenomenon has been reported for some other DNA viruses, including adenovirus, murine

cytomegalovirus, and herpes simplex virus (Anderson, M., et al., Cell 43:215-222, 1985; Burgert and Kvist, Cell 41:987-997, 1985; Heise T. M., et al., J. Exp. Med. 187:1037-1046, 1998). In the adenovirus and herpes simplex virus systems, the product of a viral gene which is dispensable for replication in vitro is sufficient to cause down-regulation of MHC class I heavy chains (Anderson, 5 M., et al., 1985, supra). The gene(s) involved in class I heavy chain down-regulation by murine cytomegalovirus have not yet been identified.

It is believed that the proteins of SEQ ID NOs:283 and 286 are members of the immunoglobulin superfamily and, as such, play a role in the immune response, cellular proteolysis, cell proliferation and differentiation, pathogen recognition, apoptosis, and other processes 10 associated with the Ig superfamily. In addition, the proteins of the invention are thought to be tightly linked to the antigen processing and presentation system in the context of peptide assembly and translocation of foreign peptides across endoplasmic reticulum and cell surface membranes as new chaperonin-like proteins associated with MHC I and TAP proteins. The weak homology (30%) with the TAP protein family is thought to indicate the specificity of the interactions of the proteins 15 of the invention with MHC proteins and/or TAP-related proteins, as described by Suling et al., supra.

Preferred polypeptides of the invention are polypeptides comprising the amino acids of SEQ ID NO:283 from position 9 to 23, 199 to 219, 205 to 285, 318 to 384, 319 to 336, 380 to 386 and from 406 to 426. Other preferred polypeptides of the invention are fragments of SEQ ID 20 NO:283 having any of the biological activities described herein.

In one embodiment, the invention relates to methods and compositions for using the protein of the invention or part thereof as a marker protein to selectively identify tissues, such as salivary glands and bone marrow tissues, which strongly express the protein of the invention. For example, the protein of the invention or part thereof may be used to synthesize specific antibodies using any 25 techniques known to those skilled in the art including those described therein. Such tissue-specific antibodies may then be used to identify tissues of unknown origin, for example, forensic samples, differentiated tumor tissue that has metastasized to foreign bodily sites, or to differentiate different tissue types in a tissue cross-section using immunochemistry.

In another embodiment, the invention relates to methods for using the protein of the 30 invention to visualize proteins and peptides involved in antigen recognition system within cells by virtue of their physical interaction with the proteins of the invention. For example, the protein may be used to detect the presence and/or the localization of MHC peptides and TAP- like proteins in a cell. The protein of the invention, and hence any interacting proteins, can be labeled using any of a number of methods, including by binding with specific antibodies or by creating a fusion protein 35 comprising the protein of the invention as well as a readily detectable moiety, such as an epitope tag, biotin, or green fluorescent protein.

In another embodiment, polynucleotide or polypeptide sequences of the invention or part thereof may be used for the diagnosis of a disorder associated with a loss of regulation of the expression of the protein of the invention, preferably, but not limited to, deficiencies of the MHC protein system. Examples of such disorders include, but are not limited to, acquired

5 immunodeficiency syndrome (AIDS), X-linked agammaglobinemia of Bruton, common variable immunodeficiency (CVI), DiGeorge's syndrome (thymic hypoplasia), thymic dysplasia, isolated IgA deficiency, severe combined immunodeficiency disease (SCID), immunodeficiency with thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi syndrome, chronic granulomatous diseases, hereditary angioneurotic edema, immunodeficiency associated with

10 Cushing's disease, Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis,

15 glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjogren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis,

20 Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, leukemias such as multiple myeloma, and lymphomas such as Hodgkin's disease; a cell proliferative disorder such as arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma,

25 melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; and an infection, such as infections by viral agents classified as adenovirus, arenavirus, bunyavirus, calicivirus, coronavirus, filovirus, hepadnavirus,

30 herpesvirus, flavivirus, orthomyxovirus, parvovirus, papovavirus, paramyxovirus, picornavirus, poxvirus, reovirus, retrovirus, rhabdovirus, and togavirus; infections by bacterial agents classified as pneumococcus, staphylococcus, streptococcus, bacillus, corynebacterium, clostridium, meningococcus, gonococcus, listeria, moraxella, kingella, haemophilus, legionella, bordetella, gram-negative enterobacterium including shigella, salmonella, and campylobacter, pseudomonas,

35 vibrio, brucella, francisella, yersinia, bartonella, norcardium, actinomyces, mycobacterium, spirochaetale, rickettsia, chlamydia, and mycoplasma; infections by fungal agents classified as aspergillus, blastomyces, dermatophytes, cryptococcus, coccidioides, malassezia, histoplasma, and

other fungal agents causing various mycoses; and infections by parasites classified as plasmodium or malaria-causing, parasitic entamoeba, leishmania, trypanosoma, toxoplasma, pneumocystis carinii, intestinal protozoa such as giardia, trichomonas, tissue nematodes such as trichinella, intestinal nematodes such as ascaris, lymphatic filarial nematodes, trematodes such as schistosoma, and cestodes such as tapeworm. To assess abnormal expression of the present protein associated with any of these disorders, the level of the present polynucleotides or polypeptides can be detected in a biological sample or cell using any standard method, including Southern or northern analysis, dot blots, other membrane-based technologies, PCR technologies, dipstick, pin, ELISA assays, and in microarrays. Any of these methods may be used for the diagnosis of disorders characterized by an alteration of expression of SEQ ID NO:283 or 286, such as the disorders mentioned above, or in assays to monitor patients being treated with SEQ ID NO:283 or 286 or agonists, antagonists, or inhibitors of SEQ ID NO:283 or 286. Antibodies useful for diagnostic purposes may be prepared, e.g., in the same manner as that described in U.S. Patent No. 6,135,941. Diagnostic assays for SEQ ID NO:283 or 286 include methods which utilize the antibody and a label to detect SEQ ID NO: 283 or 286 in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

In another embodiment, the protein of SEQ ID NO:283 or 286 or a fragment or derivative thereof may be administered to a subject to diagnose, treat or prevent an immune disorder associated with decreased expression or activity of the protein of the invention. Such disorders can include, but are not limited to, acquired immunodeficiency syndrome (AIDS), X-linked agammaglobinemia of Bruton, common variable immunodeficiency (CVI), DiGeorge's syndrome (thymic hypoplasia), thymic dysplasia, isolated IgA deficiency, severe combined immunodeficiency disease (SCID), immunodeficiency with thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi syndrome, chronic granulomatous diseases, hereditary angioneurotic edema, immunodeficiency associated with Cushing's disease, Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjogren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, leukemias

such as multiple myeloma, and lymphomas such as Hodgkin's disease. In addition, such disorders associated with decreased protein expression or activity can be treated by administering to a patient polynucleotide sequences encoding the protein of the invention, e.g. inserted in an appropriate vector. In another example, a compound that increases either the activity of the protein of the invention or their expression can be administered to a patient to treat or prevent any of the diseases mentioned above.

In a further embodiment, an antagonist of the protein of the invention may be administered to a subject to treat or prevent an immune disorder associated with increased expression or activity of the protein of SEQ ID NO:283 or 286 including, but not limited to, auto-immune diseases or graft rejection. In one aspect, an antibody which specifically binds the protein of the invention may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express the proteins of the invention, such as the salivary gland tissue or the bone marrow tissue. In addition, sense, antisense nucleotides, GSE, ribozymes, specific protein inhibitors such as antibodies or small compounds can be administered to inhibit the expression of the proteins of the invention.

In another embodiment, an antagonist of the protein of SEQ ID NO:283 may be administered to a subject to treat or prevent a cell proliferative disorder. Such disorders may include, but are not limited to, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. In one aspect, an antibody which specifically binds the protein of the invention may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express the protein of the invention. In another example, sense, antisense nucleotides, GSE, or ribozymes designed from nucleotides of the invention can be administered to inhibit the expression of the protein of the invention.

Protein of SEQ ID NO: 411 (internal designation 181-10-1-0-C9-CS)

The protein of SEQ ID NO: 411 encoded by the cDNA of SEQ ID No: 170 is highly expressed in fetal liver. The protein of the invention is homologous to peripheral benzodiazepine receptor/isoquinoline binding protein (PBR/IBP) of human, bovine and murine origin (Genbank accession numbers M36035, M64520 and L17306 respectively). The 170-amino-acid protein of SEQ ID NO: 411 is similar in size and hydropathicity to known peripheral PBR/IBP benzodiazepine receptors/isoquinoline binding proteins. Like the known peripheral benzodiazepine

receptors/isoquinoline binding proteins, the protein of the subject invention has about five potential transmembrane domains at positions 3-23, 45-65, 82-102, 105-125 and 130-150. Moreover, the protein of the invention displays a stretch of 11 amino acids (starting with V144 and ending with R154) that corresponds to a recently identified putative cholesterol recognition/interaction amino acid consensus pattern (-L/V-(X)(1-5)-Y-(X)(1-5)-R/K-) [See Li *et al*, *Endocrinology* 1998 Dec; 139 (12): 4991-7].

The peripheral benzodiazepine receptor (PBR) is a 18-kDa protein containing binding sites for benzodiazepine and is distinct from the GABA neurotransmitter receptor [Papadopoulos, V. (1993) *Endocr. Rev.* 14: 222-240]. Expression of PBR has been found in every tissue examined. However, it is most abundant in steroidogenic cells and is also found, primarily, on outer mitochondrial membranes [Anholt, R *et al.* (1986) *J. Biol. Chem.* 261:576-583]. PBR is thought to be associated with a multimeric complex composed of the 18-kDa isoquinoline binding protein and the 34-kDa pore-forming voltage dependent anion channel protein, preferentially located on the outer/inner mitochondrial membrane contact sites [McEnery, M.W. *et al. Proc. Natl. Acad. Sci. USA.* 89:3170-3174; Garnier, M. *et al.* (1994) *Mol. Pharmacol.* 45:201-211; Papadopoulos, V. *et al.* (1994) *Mol. Cel. Endocr.* 104:R5-R9]. Drug ligands of PBR, upon binding to the receptor, simulate steroid synthesis in steroidogenic cells in vitro [Papadopoulos, V *et al.* (1990) *J. Biol. Chem.* 265: 3772-3779; Barnea, E. R. *et al.* (1989) *Mol. Cell. Endocr.* 64: 155-159; Amsterdam, A. and Suh, B.S. (1991) *Endocrinology* 128: 503-510]. Likewise, *in vivo* studies showed that high affinity PBR ligands increase steroid plasma levels in hypophysectomized rats [Amri, H. *et al.* (1996) *Endocrinology* 137:5707-5718]. Further in vitro studies on isolated mitochondria provided evidence that PBR ligands, drug ligands, or the endogenous PBR ligand (the polypeptide diazepam-binding inhibitor (DBI) [Papadopoulos, V. *et al.* (1997) *Steroids* 62: 21-28]) stimulate pregnenolone formation by increasing the rate of cholesterol transfer from the outer to the inner mitochondrial membrane [for review, see Culty, M. *et al.* (1999) *Journal of Steroid Biochemistry and Molecular Biology* 69: 123-130].

Based on the amino acid sequence of the 18-kDa PBR, a three dimensional model was developed [Papadopoulos, V. (1996) In: *The Leydig Cell*. Payne, A. H. *et al.* (eds) Cache River Press, IL, pp 596-628]. This model was shown to accommodate a cholesterol molecule and function as a channel, supporting the role of PBR in cholesterol transport. The role of PBR in steroidogenesis was also demonstrated by observing that PBR negative cells generated by homologous recombination failed to produce steroids [Papadopoulos, V. *et al* (1997) *J. Biol. Chem.* 272: 32129-32135]. Further, cholesterol transport experiments in bacteria expressing the 18-kDa PBR protein provided definitive evidence for a function as a cholesterol channel/transporter [Papadopoulos, V. *et al.* (1997) *supra*].

In addition to its role in mediating cholesterol movement across membranes, PBR has been implicated in several other physiological functions, including cell growth and differentiation,

chemotaxis, mitochondrial physiology, porphyrin and heme biosynthesis, immune response, anion transport and GABAergic regulation of CNS. [for review, see Gavish, M. *et al.* (1999) *Pharmaceutical Reviews* 51: 629-650; Beurdeley-Thomas, A. *et al.* (2000) *Journal of Neuro-Oncology* 46: 45-56]. Also, a recent report also indicates that PBR agonists are potent anti-apoptotic compounds. These findings suggest that this effect may represent a major function for this receptor (Bono, F. *et al.* (1999) *Biochemical and Biophysical Research Communications* 265:457-461].

It appears that PBR is associated with stress and anxiety disorders. It has been suggested that PBRs play a role in the regulation of several stress systems such as the HPA axis, the sympathetic nervous system, the renin-angiotensin axis, and the neuroendocrine axis. In these systems, acute stress typically leads to increases in PBR density, whereas chronic stress typically leads to decreases in PBR density. Furthermore, in Generalized Anxiety Disorder (GAD), Panic Disorder (PD), Generalized Social Phobia (GSP), and Post-Traumatic Stress Disorders (PTSD), PBR density is typically decreased in platelets.

In the brain, where PBRs are associated with glial cells, PBRs are increased in specific brain areas in neurodegenerative disorders and also after neurotoxic and traumatic-ischemic brain damage [for review, see Gavish, M. *et al.* (1999) *supra*]. The literature also reports a decrease in peripheral-type benzodiazepine receptors in postmortems of chronic schizophrenics, suggesting that the decreased density of PBRs in the brain may be involved in the pathophysiology of schizophrenia. Increased levels of PBR in autopsied brain tissue from PSE patients (Portal-Systemic Encephalopathy patients) have been reported, thus supporting the theory that activation of PBR contributes to the pathogenesis characteristic of portal-systemic encephalopathy (PSE) in the central nervous system [Kurumaji, A. *et al.* (1997) *J. Neural Transm* 104:1361-1370; Butterworth R. F. (2000) *Neurochemistry International* 36: 411-416].

In addition to its involvement in the neurological disorders discussed *supra*, PBR has been implicated in the regulation of tumor cell proliferation [for review, see Gavish, M. *et al.* (1999) *supra*; Beurdeley-Thomas, A. *et al.* (2000) *supra*; Hardwick, M. (1999) *Cancer Research* 59:831-842; Venturini, I. *et al.* (1998) *Life Sci* 63:1269-80; Carmel I *et al.* (1999) *Biochem Pharmacol* 58: 273-8]. The invasiveness and metastatic ability of human breast tumor cells is proportional to the level of PBR expressed. Further, PBR has been proposed to be used as a tool/marker for detection, diagnosis, prognosis and treatment of cancer [WO 99/49316, hereby incorporated by reference in its entirety].

Many ligands have been described that bind to peripheral benzodiazepine receptor with various affinities. Some benzodiazepines, Ro 5-4864 [4-chlorodiazepam], diazepam and structurally related compounds, are potent and selective PBR ligands. Exogenous ligands also include 2-phenylquinoline carboxamides (PK11195 series), imidazo [1,2-a]pyridine-3-acetamides (Alpidem series) and pyridazine derivatives. Some endogenous compounds, including porphyrins and

diazepam binding inhibitor (DBI), bind to PBR with nanomolar and micromolar affinity [for review, see Gavish, M. *et al.* (1999) *supra*; Beurdeley-Thomas, A. *et al.* (2000) *supra*].

The protein of SEQ ID NO: 411 is a novel peripheral-type benzodiazepine receptor. As such, it serves a channel function that mediates cholesterol movement across membranes, play a role in steroidogenesis, cell growth and differentiation, chemotaxis, mitochondrial physiology, protection against apoptosis, porphyrin and heme biosynthesis, immune response, anion transport and GABAergic regulation of CNS.

In one embodiment, a preferred polypeptide of the invention comprises the amino acids of SEQ ID NO: 411 from position 144 to 154. In another embodiment, the subject invention provides a polypeptide comprising the sequence of SEQ ID NO: 411. Other preferred polypeptides of the invention include biologically active fragments of SEQ ID NO: 411. Biologically active fragments of the protein of SEQ ID NO: 411 have any of the biological activities described herein which are associated with the PBR. In another embodiment, the polypeptide of the invention is encoded by clone 181-10-1-0-C9-CS.

One aspect of the subject invention provides compositions and methods using the protein of the invention, or biologically active fragments thereof, for the development, identification, and/or selection of agents capable of modulating the expression or activity of the protein of the invention.

Agents which modulate the activity of the PBR/IBP of the subject invention include, but are not limited to, antisense oligonucleotides, ribozymes, drugs, and antibodies. These agents may be made and used according to methods well known in the art. Also, the protein of the invention, or biologically active fragments thereof, may be used in screening assays for therapeutic compounds. A variety of drug screening techniques may be employed. In this aspect of the invention, the protein or biologically active fragment thereof, may be free in solution, affixed to a solid support, recombinantly expressed on, or chemically attached to, a cell surface, or located intracellularly. The formation of binding complexes, between the protein of the invention, or biologically active fragments thereof, and the compound being tested, may then be measured.

In one embodiment, the subject method utilizes eukaryotic or prokaryotic host cells which are stably transformed with recombinant nucleic acids expressing the PBR/IBP polypeptide or biologically active fragments thereof. The transformed cells may be viable or fixed. Drugs or compounds which are candidates for the modulation of the PBR/IBP, or biologically active fragments thereof, are screened against such transformed cells in binding assays well known to those skilled in the art. Alternatively, assays such as those taught in Geysen H. N., WO Application 84/03564, published on Sep. 13, 1984, and incorporated herein by reference in its entirety, may be used to screen for peptide compounds which demonstrate binding affinity for, or the ability to modulate, the PBR/IBP, or biologically active fragments thereof. In another embodiment, competitive drug screening assays using neutralizing antibodies specifically compete with a test

compound for binding to the PBR/IBP protein of the invention, or biologically active fragments thereof.

Another embodiment of the subject invention provides compositions and methods of selectively modulating the expression or activity of the protein of the invention. Modulation of the PBR/IBP would allow for the successful treatment and/or management of diseases or biochemical abnormalities associated with the PBR or PBR/IBP. Antagonists, able to reduce or inhibit the expression or the activity of the protein of the invention, would be useful in the treatment of diseases associated with elevated levels of the PBR/IBP, increased cell proliferation, or increased cholesterol transport. Thus, the subject invention provides methods for treating a variety of diseases or disorders, including, but not limited to, cancers, especially liver cancer, and portal-systemic encephalopathy.

Alternatively, the subject invention provides methods of treating diseases or disorders associated with decreased levels of the protein of the PBR/IBP. Thus, the subject invention provides methods of treating diseases including, and not limited to, schizophrenia, chronic stress, GAD, PD, GSP and PTSD. Other diseases which may be treated by agonists of the PBR/IBP of the subject invention include those diseases associated with decreases in cell proliferation, e.g. developmental retardation.

Furthermore, because the PBR/IBP of the subject invention is also able to transport cholesterol into cells, the subject invention may also be used to increase cholesterol transport into cells. Diseases associated with cholesterol transport deficiencies include lipoidal adrenal hyperplasia, and diseases where there is a requirement for increased production of compounds requiring cholesterol such as myelin and myelination, such as Alzheimer's disease, spinal chord injury, and brain development neuropathy [Snipes, G. and Suter, U. (1997) Cholesterol and Myelin. In: Subcellular Biochemistry, Robert Bittman (ed.), vol. 28, pp. 173-204, Plenum Press, New York]. The methods of treating disorders associated with decreased levels of PBR/IBP may be practiced by introducing agonists which stimulate the expression or the activity of the protein of the invention.

In one embodiment, methods of increasing the levels of PBR/IBP in tissues or cell types may be practiced by utilizing nucleic acids encoding the protein of the subject invention, or biologically active fragments thereof, to introduce biologically active polypeptide into targeted cell types. Vectors useful in such methods are known to those skilled in the art as are methods of introducing such nucleic acids into target tissues.

Agents which stimulate or inhibit the activity of the protein of the invention include but are not limited to agonist and antagonist drugs respectively. These drugs can be obtained using any of a variety of drug screening techniques as discussed above.

Antagonists of the PBR/IBP encoded by SEQ ID NO: 170 include agents which decrease the levels of expressed mRNA encoding the protein of SEQ ID NO: 411. These include, but are not limited to, RNAi, one or more ribozymes capable of digesting the protein of the invention mRNA,

or antisense oligonucleotides capable of hybridizing to mRNA encoding the PBR/IBP of SEQ ID NO: 411. Antisense oligonucleotides can be administrated as DNA, as DNA entrapped in proteoliposomes containing viral envelope receptor proteins [Kanoda, Y. *et al.* (1989) *Science* 243: 375] or as part of a vector which can be expressed in the target cell and provide antisense DNA or RNA. Vectors which are expressed in particular cell types are known in the art. Alternatively, the DNA can be injected along with a carrier. A carrier can be a protein such as a cytokine, for example interleukin 2, or polylysine-glycoprotein carriers. Carrier proteins, vectors, and methods of making and using polylysine carrier systems are known in the art. Alternatively, nucleic acid encoding antisense molecules may be coated onto gold beads and introduced into the skin with, for example, a gene gun [Ulmer, J.B. *et al.* (1993) *Science* 259:1745].

Antibodies, or other polypeptides, capable of reducing or inhibiting the activity of PBR/IBP may be provided as in isolated and substantially purified form. Alternatively, antibodies or other polypeptides capable of inhibiting or reducing the activity of the PBR/IBP protein, may be recombinantly expressed in the target cell to provide a modulating effect. In addition, compounds which inhibit or reduce the activity of the PBR/IBP protein of the subject invention may be incorporated into biodegradable polymers being implanted in the vicinity of where drug delivery is desired. For example, biodegradable polymers may be implanted at the site of a tumor or, alternatively, biodegradable polymers containing antagonists/agonists may be implanted to slowly release the compounds systemically. Biodegradable polymers, and their use, are known to those of skill in the art (see, for example, Brem *et al.* (1991) *J. Neurosurg.* 74:441-446).

In another embodiment, the invention provides methods and compositions for detecting the level of expression of the mRNA of the protein of the invention. Quantification of mRNA levels of the PBR/IBP protein of the invention may be useful for the diagnosis or prognosis of diseases associated with an altered expression of the protein of the invention. Assays for the detection and quantification of the mRNA of the protein of the invention are well known in the art (see, for example, Maniatis, Fritsch and Sambrook, *Molecular Cloning; A Laboratory Manual* (1982), or *Current Protocols in Molecular Biology*, Ausubel, F.M. *et al.* (Eds), Wiley & Sons, Inc.).

Polynucleotides probes or primers for the detection of the mRNA of the protein of SEQ ID NO: 411 can be designed from the cDNA of SEQ ID NO: 170. Methods for designing probes and primers are known in the art. In another embodiment, the subject invention provides diagnostic kits for the detection of the mRNA of the protein of the invention in cells. The kit comprises a package having one or more containers of oligonucleotide primers for detection of the protein of the invention in PCR assays or one or more containers of polynucleotide probes for the detection of the mRNA of the protein of the invention by in situ hybridization or Northern analysis. Kits may, optionally, include containers of various reagents used in various hybridization assays. The kit may also, optionally, contain one or more of the following items: polymerization enzymes, buffers, instructions, controls, or detection labels. Kits may also, optionally, include containers of reagents

mixed together in suitable proportions for performing the hybridization assay methods in accordance with the invention. Reagent containers preferably contain reagents in unit quantities that obviate measuring steps when performing the subject methods.

In another embodiment, the invention relates to methods and compositions for detecting and
5 quantifying the level of the protein of the invention present in a particular biological sample. These methods are useful for the diagnosis or prognosis of diseases associated with an altered levels of the protein of the invention. Diagnostic assays to detect the protein of the invention may comprise a biopsy, in situ assay of cells from organ or tissue sections, or an aspirate of cells from a tumor or normal tissue. In addition, assays may be conducted upon cellular extracts from organs, tissues,
10 cells, urine, or serum or blood or any other body fluid or extract.

Assays for the quantification of the PBR/IBP of SEQ ID NO: 411 may be performed according to methods well known in the art. Typically, these assays comprise contacting the sample with a ligand of the protein of the invention or an antibody (polyclonal or monoclonal) which recognizes the protein of the invention or a fragment thereof, and detecting the complex formed
15 between the protein of the invention present in the sample and the ligand or antibody. Fragments of the ligands and antibodies may also be used in the binding assays, provided these fragments are capable of specifically interacting with the BRP/IRP of the subject invention. Further, the ligands and antibodies which bind to the BRP/IRP of the invention may be labeled according to methods known in the art. Labels which are useful in the subject invention include, but are not limited to,
20 enzymes labels, radioisotopic labels, paramagnetic labels, and chemiluminescent labels. Typical techniques are described by Kennedy, J. H., *et al.* (1976) *Clin. Chim. Acta* 70:1-31; and Schurs, A. H. *et al.* (1977) *Clin. Chim. Acta* 81: 1-40.

The subject invention also provides methods and compositions for the identification of metastatic tumor masses. In this aspect of the invention, the polypeptides and antibodies which
25 bind the polypeptides of the invention may be used as a marker for the identification of the metastatic tumor mass. Metastatic tumors which originated from the liver may overexpress the PBR/IBP of SEQ ID NO: 411, whereas newly forming tumors, or those originating from other tissues are not expected to bear the PBR/IBP of SEQ ID NO: 411.

Protein of SEQ ID NO: 397 (internal designation 160-28-4-0-C4-CS).

30 The protein of SEQ ID NO: 397, encoded by the cDNA of SEQ ID NO: 156 (clone 160-28-4-0-C4-CS), exhibits homology to the ADP-ribosylation factors (ARF) family of proteins. The ARF family includes ADP-ribosylation factors (ARFs) and ARF-like proteins (ARLs); the ARF family of proteins is one family of the Ras superfamily. Proteins belonging to the Ras superfamily have molecular weights of 18-30 kDa and function in a variety of cellular processes including, but
35 not limited to, signaling, growth, immunity, and protein transport.

ARFs are monomeric GTP-binding proteins, related structurally to both G protein alpha-subunits and Ras proteins. ARF family members share more than 60% sequence identity, appear to be ubiquitous in eukaryotes, and are evolutionarily highly conserved throughout. Immunologically, they have been localized to the Golgi apparatus of several types of cells (Stearns et al. Proc. Natl. Acad. Sci. (USA) 87:1238-1242 (1990)). ARF proteins enhance the ADP-ribosyltransferase activity of cholera toxin as an allosteric activator (Noda et al. Biochim. Biophys. Acta 1034: 195-199 (1990)). ARFs have also been shown to act as regulatory molecules, or "switches", for linking two processes (e.g., the process of vesicle fission from a donor compartment and fusion with an acceptor compartment (Rothman, J. E. and Wieland, F. T. Science 272: 227-234 (1996)). ARF family members fall into three classes, classes I-III, according to their size and sequence homology. Class I comprises ARF1, ARF2, and ARF3; Class II comprises ARF4 and ARF5; and Class III comprises ARF6.

The classes occupy different subcellular locations and have been implicated in different transport pathways. Class I ARFs localize to the Golgi where they are involved in the regulation of ER-Golgi and intra-Golgi transport. Class I ARFs are also involved in the recruitment of cytosolic coat proteins to Golgi membranes during the formation of transport vesicles. Class III (e.g., ARF6) localizes to a tubulovesicular compartment, secretory granules, and the plasma membrane, where it is involved in regulated secretion and recycling. Class II ARFs appear to be cytosolic, but their role has not been elucidated. (Radhakrishna, H. and Donaldson, J. G. J. Cell Biol. 139: 49-61(1997)).

ARF function, in general, is regulated by a GDP-GTP cycle. For example, ARF1 is cytosolic in the GDP bound state, but is associated with membranes when in the GTP bound state. A guanine nucleotide exchange factor (GEF) in the donor compartment recruits ARF1 to the membrane. At the membrane, GTP-ARF1 recruits coat proteins, which assemble together into spherical coats, budding off vesicles in the process. After budding, hydrolysis of bound GTP causes ARF1 to dissociate from the membrane. ARF1 dissociation causes the coat to become unstable and dissociate as well. (Rothman, supra.)

Members of the ARF multigene family, when expressed as recombinant proteins in *E. coli*, display different phospholipid and detergent requirements (Price, et al. J. Biol. Chem. 267: 17766-17772 (1992)). Some lipids and/or detergents, e.g., SDS, cardiolipin, dimyristoylphosphatidylcholine (DMPC)/cholate, enhance ARF activities (Bobak, et al. Biochemistry 29:855-861 (1990); Noda, et al. Biochim. Biophys. Acta 1034: 195-199 (1990); Tsai, et al. J. Biol. Chem. 263:1768-1772 (1988)). ARFs also activate phospholipase D (PLD), a membrane-bound enzyme implicated as an effector of several growth factors (Boman, A. L. and Kahn, R. A. Trends Biochem. Sci. 20: 147-150 (1995). PLD1 has been shown to be activated by a variety of G-protein regulators, for example, PKC (protein kinase C) and ADP-ribosylation factor (ARF). PKC and ARFs may regulate G-proteins either individually or together in a synergistic manner. Recently the role of ARFs in microtubules formation has also been demonstrated. ADP-

ribosylation of tubulin almost completely blocked self-assembly of this protein in brain (Terashima M. et al; J.Nutr Sci Vitaminol 45: 393-400 (1999)).

In general, differences in the various ARF sequences are concentrated in the amino-terminal regions and the carboxyl portions of the proteins. Only three of 17 amino acids in the amino termini have shown to be identical among ARFs, and four amino acids in this region of ARFs 1-5 are missing in ARF 6 (Tsuchiya, et al. J. Biol. Chem. 266: 2772-2777 (1991)). It was reported (Kahn, et al. J. Biol. Chem. 267:13039-13046 (1992)) that the amino-terminal regions of ARF proteins form an alpha-helix and that this domain is required for membrane targeting, interaction with lipid, and ARF activity.

Schliefer et al., (J. Biol. Chem. 257: 20-23 (1991)) have described a protein distinctly larger than ARF that possessed ARF-like activity. ARF-like proteins, or ARLs, have been found in different species. Some of ARLs appear to lack ADP-ribosyltransferase-enhancing activity; ARLs may differ in GTP-binding requirements and GTPase activity as compared to various ARF isoforms. For example, ARP, a mammalian ARL, is 33-39% identical to members of the ARF family; ARP, however, differs from other ARF family proteins by virtue of its ability to hydrolyze bound GTP in the absence of other proteins. ARP protein, unlike ARFs, is typically associated with plasma membrane instead of the cytosol (Schurmann, A. J. Biol. Chem. 270, 30657-30663 (1995)).

ARF family members have been implicated in several disease processes, such as Lowe's syndrome, an X-linked disorder characterized by congenital cataracts, renal tubular dysfunction and neurological deficits. These disorders may be due to an inability to recruit ARF to the Golgi membrane (Suchy, S. F. et al. Hum. Mol. Genet. 4: 2245-2250 (1995), Londono I. et al. Kidney Int. 55: 1407-1416 (1999)). It has also been suggested that regulation of ARF is also involved in cystic fibrosis, Dent's disease, diabetes, and autosomal dominant polycystic kidney disease (Marshansky, V., et al. Electrophoresis 18: 2661-2676 (1997)).

The new human ARF-related protein of SEQ ID NO:397, encoded by clone 160-28-4-0-C4-CS in one embodiment, and the related polynucleotides, provide new compositions which are useful in the diagnosis, treatment, and prevention of secretory, exocytosis, endocytosis and another "sorting disorders."

The subject invention provides a polypeptide comprising the amino acid sequence of SEQ ID NO: 397 or clone 160-28-4-0-C4-CS, or biologically active fragments thereof. The intact protein of interest is 173 amino acids in length, has an ARF family amino acid motif (Pfam), and has ATP/GTP-binding site motif A P-loop (PS00017). The protein of SEQ ID NO: 397 or clone 160-28-4-0-C4-CS also has chemical and structural similarity with human ARL1 (P40616), ARD-1 (R66033) and ARF6 (GI 178989) (31%, 31% and 27% identity, respectively). The amino acid length of SEQ ID NO: 397 is similar to those of the aforementioned ARFs. Biologically active fragments of SEQ ID NO: 397 have one or more of the biological activities typically associated the full length protein. In one embodiment, the protein is encoded by clone 160-28-4-0-C4-CS

The invention also provides variants of the protein of SEQ ID NO: 397 or clone 160-28-4-0-C4-CS. The variants have at least about 80%, more preferably at least about 90%, and most preferably at least about 95% amino acid sequence identity to the amino acid sequence of SEQ ID NO: 397 or clone 160-28-4-0-C4-CS. Variants according to the subject invention have at least one functional and/or structural characteristic of ARFs. The invention also provides biologically active fragments of the variant proteins.

The invention includes those polynucleotides encoding the protein of SEQ ID NO: 397 or clone 160-28-4-0-C4-CS, variants of SEQ ID NO: 397 or clone 160-28-4-0-C4-CS, and biologically active fragments of both the protein of SEQ ID NO: 397 or clone 160-28-4-0-C4-CS and variants thereof. As is apparent to those skilled in the art, a variety of different DNA sequences can encode the amino acid sequence of the proteins, variants, and biologically active fragments of said proteins and variants. It is well within the skill of a person trained in the art to create these alternative DNA sequences encoding proteins having the same, or essentially the same, amino acid sequence. These variant DNA sequences are also within the scope of the subject invention. As used herein, reference to "essentially the same" sequence refers to sequences that have amino acid substitutions, deletions, additions, or insertions that do not materially affect biological activity.

The subject invention provides method of treating cytoskeletal, secretory, and inflammatory disorders/conditions comprising the administration of therapeutically effective amounts of a composition comprising the protein of SEQ ID NO: 397 or clone 160-28-4-0-C4-CS. These methods can also be practiced using variants of SEQ ID NO: 397 or clone 160-28-4-0-C4-CS, or biologically active fragments of either SEQ ID NO: 397 or clone 160-28-4-0-C4-CS, or variants of SEQ ID NO: 397 or clone 160-28-4-0-C4-CS. Disorders/conditions which can be treated by the subject invention include, but are not limited to, prostate cancer, brain and another tumors, Lowe's syndrome, glomerulonephritis, chronic glomerulonephritis, tubulointerstitial nephritis, inherited X-linked nephrogenic diabetes insipidus, autosomal dominant polycystic kidney disease (ADPKD), herpes gestationis, dermatitis herpetiformis, lupus erythematosus, Crohn's disease, irritable bowel syndrome and Addison's disease; secretory/endocytotic disorders such as cystic fibrosis, glucose-galactose malabsorption syndrome, hypercholesterolemia, hyper- and hypoglycemia, Grave's disease, goiter, and Cushing's disease; conditions associated with abnormal vesicle trafficking, including acquired immunodeficiency syndrome (AIDS); allergies including hay fever, asthma, and urticaria (hives); autoimmune hemolytic anemia; multiple sclerosis; myasthenia gravis; rheumatoid and osteoarthritis; Chediak-Higashi and Sjogren's syndromes; toxic shock syndrome; traumatic tissue damage; viral, bacterial, fungal, helminthic, and protozoal infections.

In another embodiment, a vector capable of expressing the protein of SEQ ID NO: 397 or clone 160-28-4-0-C4-CS, or biologically active fragments thereof, can be administered to a subject to treat or prevent disorders including, but not limited to, those described above. Alternatively, the vector can encode a variant, or biologically active fragment of the variant protein. Multiple vectors

encoding any combination of SEQ ID NO: 397 or clone 160-28-4-0-C4-CS, variants, and/or biologically active fragments of SEQ ID NO: 397 or clone 160-28-4-0-C4-CS and/or variants can be administered to a subject.

In a further embodiment, a pharmaceutical composition comprising a substantially purified protein of SEQ ID NO: 397 or clone 160-28-4-0-C4-CS (and/or biologically active fragments thereof), in conjunction with a suitable pharmaceutical carrier, can be administered to a subject to treat or prevent the above mentioned disorders. Alternatively, a pharmaceutical composition comprising a substantially purified variant protein of SEQ ID NO: 397 or clone 160-28-4-0-C4-CS (and/or biologically active fragments thereof), in conjunction with a suitable pharmaceutical carrier, can be administered in the aforementioned therapeutic regimens. As would be apparent to the skilled artisan, any therapeutically effective combination of the protein encoded by SEQ ID NO: 397 or clone 160-28-4-0-C4-CS (and/or biologically active fragments thereof) and variants of SEQ ID NO: 397 or clone 160-28-4-0-C4-CS (and/or biologically active fragments thereof), in conjunction with a suitable pharmaceutical carrier can be used in the aforementioned therapeutic regimens.

ARFs are known to be involved in regulated transport of vesicles. Therefore, in another embodiment, the protein of SEQ ID No: 397 or clone 160-28-4-0-C4-CS, variants, and/or biologically active fragments of said proteins and/or variants can be used as a component of drug delivery vehicles such as colloids or liposomes. The protein of SEQ ID NO: 397 or clone 160-28-4-0-C4-CS, variants, and/or biologically active fragments of said proteins and/or variants can be incorporated into the lipid membranes of liposomes and can serve as specific targeting agents. The methods of design of such drug delivery systems is known by those skilled in the art and can be practiced according to conventional pharmaceutical principles (Smith H.J. Introduction to the principles of drug design and action, 3rd ed. (1998); Chien Y.W. Novel Drug Delivery systems, 2nd ed. (1992); Storm G. et al J.Liposome Res. 4: 641-666 (1994); and Crommelin D.J.A. et al. Adv. Drug Delivery Rev. 17 : 49-60 (1995)).

In another embodiment of the invention, the polynucleotides encoding the protein of SEQ ID NO: 397 or clone 160-28-4-0-C4-CS can be used for therapeutic purposes. Polynucleotides encoding fragments of the protein of SEQ ID NO: 397 or clone 160-28-4-0-C4-CS can also be used in therapeutic regimens. In one aspect, the complement of the polynucleotide encoding the protein of SEQ ID NO: 397 or clone 160-28-4-0-C4-CS can be used in situations in which it would be desirable to block the transcription of the mRNA. Modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5', or regulatory regions of the gene encoding the protein of interest. Such technology is now well known in the art, and sense or antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding the

protein of interest. Methods of treatment utilizing antisense technology are also well known to those skilled in the art.

Another embodiment of the invention provides methods of assessing PLD modulation by using ARF properties of the protein of interest.

5 In another embodiment, antibodies which specifically bind the protein of SEQ ID NO: 397 or clone 160-28-4-0-C4-CS can be used for the diagnosis of disorders characterized by expression of the protein, or in assays to monitor patients being treated with the protein of interest. Methods of making both polyclonal and monoclonal antibodies are well-known in the art. Diagnostic assays which can be used in this aspect of the invention include, and are not limited to, ELISAs, RIAs, and
10 FACS, and are well known in the art. These assays also provide a basis for diagnosing or identifying altered or abnormal levels of SEQ ID NO:397 or the polypeptides encoded by the human cDNA of clone 160-28-4-0-C4-CS expression as compared to normal individuals. These screening methods are, likewise, well known to the skilled artisan.

In another embodiment of the invention, the protein of interest, its catalytic or immunogenic
15 fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening can be free in solution, affixed to a solid support, recombinantly expressed on, or chemically attached to, a cell surface, or located intracellularly. The formation of binding complexes between the protein of interest and the agent being tested can be measured by methods well known to those skilled in the
20 art. Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, *e.g.*, Geysen, et al. (1984) PCT application WO84/03564.)

In another embodiment of the invention, the polynucleotides encoding the protein of interest can be used for diagnostic purposes. The polynucleotides can be used to detect and quantify gene
25 expression in biopsied tissues in which expression of the protein of interest can be correlated with a disease or condition. Such diagnostic assays are well known in the art and can be used to monitor regulation of the protein of interest levels during therapeutic intervention and/or to determine absence, presence, and excess expression of the protein of interest. Examples of such conditions and disorders have been provided supra. The polynucleotide sequences encoding the protein of
30 interest can be used, for example, in Southern or Northern analyses, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and ELISA assays; and in microarrays utilizing fluids or tissues from patients to detect altered expression of the protein of SEQ ID NO:397 or clone 160-28-4-0-C4-CS. Such qualitative or quantitative methods are well known in the art.

35 In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein can be used as targets in a microarray. The microarray can be used to monitor the expression level of large numbers of genes simultaneously and to

identify genetic variants, mutations, and polymorphisms. This information can be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, and to develop and monitor the activities of therapeutic agents. Microarrays can be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T. M. et al. (1995) U.S. Pat. No. 5,474,796; Heller, R. A. et al. (1997) Proc. Natl. Acad. Sci. 94: 2150-2155; and Heller, M. J. et al. (1997) U.S. Pat. No. 5,605,662.)

Another embodiment of the subject invention provides nucleic acid sequences encoding the protein of interest which can be extended utilizing a partial nucleotide sequence and various PCR-based methods. This aspect of the invention provides methods for the detection of upstream sequences, such as promoters and regulatory elements. Methods of practicing this aspect of the invention are also well known in the art.

In other embodiments of the disclosed therapeutic regimens, any of the proteins, variants, biologically active fragments, antibodies, complementary sequences, or vectors of the invention can be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy can be made by one of ordinary skill in the art. The combination of therapeutic agents can act synergistically to effect the treatment or prevention of the various disorders described above. In particular, purified protein can be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind the protein of interest. Neutralizing antibodies especially preferred for therapeutic use.

Protein of SEQ ID NO: 287 (internal designation 174-5-3-0-H7-CS)

The protein of SEQ ID NO: 287, encoded by human cDNA of SEQ ID NO: 46 (clone 174-5-3-0-H7-CS), is highly homologous (more than 99% identity in amino acids) to the human protein encoded by the CLN8 gene listed in Genbank under accession number AF123757. The two proteins differ by two conservative amino-acid substitutions (alanine for valine at position 155 and serine for asparagine at position 225). In addition, the protein encoded by 174-5-3-0-H7-CS contains seven transmembrane domains. These domains are located at amino acids 25-45, 71-91, 100-120, 133-153, 160-180, 205-225, and 228-248 as predicted by the software TopPred II (Claros and von Heijne, *CABIOS applic. Notes*, 10:685-686 (1994)). The protein encoded by SEQ ID NO: 287 also exhibits a signal peptide at positions 1-50 and a retention signal KKRP from positions 283 to 286.

CLN8 was identified recently by positional cloning (Ranta et al., *Nat Genet.* 1999 Oct.;23(2):233-6). CLN8 encodes a 286 amino-acid putative transmembrane protein with no homology to previously known proteins. A naturally-occurring missense mutation in codon 24 (R24G at the border of the first putative transmembrane domain) is the molecular basis for EPMR ("progressive epilepsy with mental retardation", MIM 600143). EPMR, also called Northern Epilepsy, is an autosomal recessive disorder characterized by normal early development, onset of

generalized tonic-clonic seizures between the ages of 5 and 10 years, and subsequent progressive mental retardation. Neuropathological findings have shown that EPMR is a new member of the neuronal ceroid lipofuscinosis (NCL) group of neurodegenerative disorders. The NCLs are a genetically heterogeneous group of progressive neurodegenerative disorders characterized by the accumulation of autofluorescent lipopigment in various tissues. CLN8 is the eighth gene to be linked to the NCL group of neuro-degenerative disorders.

Subsequently, the homologous mouse gene (Cln8) was sequenced (82% nucleotide identity with the human gene) and localized to the region of the mouse genome linked to motor neuron degeneration, mouse *mnd*. *Mnd* is a naturally-occurring mouse mutant with intracellular autofluorescent inclusions similar to those seen in EPMR. A mutation in *mnd* mouse DNA was identified, indicating that *mnd* is a murine ortholog for CLN8 (Ranta et al., Nat Genet. 1999 Oct;23(2):233-6), and that mice containing mutations in *Cln8* represent a murine model for NCL disorders.

Recent experimental evidence has confirmed the transmembrane nature of the CLN8 protein (Lonka L et al., Hum Mol Genet. 2000 Jul 1;9(11):1691-7). CLN8 resides in the endoplasmic reticulum (ER) and recycles between the ER and the ER-Golgi intermediate compartment (ERGIC) via a KKXX ER-retrieval motif at its C-terminus (KKRP, amino-acids 283-286). This motif is recognized and bound by COPI, a vesicle-coating protein found in retrograde vesicles delivering cargo from the *cis* Golgi to the ER. The 30kD CLN8 protein is not processed during its maturation (in particular it is not N-glycosylated). The EPMR-associated R24G mutation does not alter cellular localization in humans.

The subject invention provides a polypeptide encoded by SEQ ID NO: 287 and biologically active fragments of said polypeptide. Compositions comprising polypeptides and pharmaceutically acceptable carriers are likewise provided. Preferred polypeptides, and biologically active fragments thereof, have any of the biological activities or domains/motifs described herein and/or contain the amino acids of positions 155 and 225, 283 to 286. In one embodiment, the protein/polypeptide of SEQ ID NO: 287 is encoded by clone 174-5-3-0-H7-CS.

The ER/ERGIC cellular localization of protein of this invention can be used to target compounds to the ER/ERGIC. This targeting can be observed using any of the techniques known to those skilled in the art including those described in (Lonka L et al., Hum Mol Genet. 2000 Jul 1;9(11):1691-7). In this aspect of the invention, the protein of SEQ ID NO: 287, or biologically active fragments thereof can be used to target liposomes, vesicles, or colloids to the ER/ERGIC compartment where active agents can be delivered. Methods of making and using targeted liposomes are well known in the art.

In another embodiment, liposomes comprising the protein of SEQ ID NO: 287 can contain a second targeting agent for the specific selection of a target cell. The second targeting agent can be selected for its ability to specifically target a cell or tissue. Thus, the second targeting agent can be

specific for tumor markers, such as HER2. Alternatively, markers associated with specific cell types can be used (e.g., CD34, CD4, CD8, etc.). In a preferred embodiment, the second targeting agent is an antibody. Active agents include, but are not limited to, chemotherapeutic agents protein cross-linking agents, inhibitors of protein synthesis, anti-bacterial agents (e.g., antibiotics), antiviral agents, and/or anti-parasitic agents. The ability to bind the COPI coatomer can be assayed as described in (Cosson P, Letourneur F, Science. 1994 Mar 18;263(5153):1629-31).

In another embodiment, the present invention provides methods of, and compositions for, identifying specific cellular compartments, such as the ER, ERGIC, and retrograde transport vesicles. This embodiment provides antibodies which specifically bind the protein of SEQ ID NO: 287, or biologically active fragments thereof, which are labeled with detectable markers, such as gold particles, enzymes, radioisotopes, or paramagnetic labels. ER, ERGIC, and retrograde transport vesicles can be identified in samples according to well-known immuno-diagnostic protocols. The antibodies, either monoclonal or polyclonal, can be made according to well-known methods. In a preferred embodiment, the antibodies bind to ER retention signal.

In another embodiment, the protein of the invention or part thereof can be used as a reagent for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, asthma, pulmonary edema, atherosclerosis, restenosis, stroke potential, thrombosis and hypertension. Similarly, the protein of the invention, or biologically active fragments thereof, and antibodies thereto can provide immunological probes for differential identification of the tissue(s) or cell type(s). In a number of disorders listed above, particularly of the pulmonary and cardiovascular systems, expression of this protein at significantly higher or lower levels can be routinely detected in certain tissues or cell types (e. g., vascular tissues, cancerous and wounded tissues) or bodily fluids (e. g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Indeed, the 80 first amino-acids of the protein of the invention are identical to two polypeptides claimed in Patent WO 99/35158, hereby incorporated by reference in its entirety (SEQ ID NO:98 and SEQ ID NO:162 corresponding to Geneseq accession numbers Y38413/Y38428 and Y38492) are over-expressed in pulmonary and endothelial tissues.

The tissue distribution in pulmonary and endothelial tissues indicates that the protein product described in WO 99/35158 is useful for the treatment and diagnosis of cardiovascular and respiratory or pulmonary disorders such as asthma, pulmonary edema, pneumonia, atherosclerosis, restenosis, stroke, angina, thrombosis hypertension, inflammation, and wound healing. Those conditions can be diagnosed by determining the amount of the protein of the invention in a sample. Thus, antibodies raised against the protein of SEQ ID NO: 287, or an immunogenic fragment of the

protein can be used in diagnostic, prognostic, or screening assays such as those taught in WO 99/35158.

Protein of SEQ ID No. 270 (internal designation 116-119-3-0-H5-CS)

The protein of SEQ ID NO: 270 encoded by the extended cDNA SEQ ID NO: 29 is
 5 homologous to the human mitochondrial ATP synthase f subunit or ATPK (E.C. 3.6.1.34) (Swissprot accession number P56134) and is overexpressed in fetal kidney.

The protein of SEQ ID NO: 270, composed of 88 amino acid residues, contains 1
 transmembrane segment (position 1 to 55) predicted by the software TopPred II (Claros and von
 Heijne, *CABIOS applic. Notes*, 10 :685-686 (1994). BLAST results show that 100% homology is found
 10 between amino acids 5 to 88 of the protein of the invention and amino acids 10 to 93 of human ATP
 synthase f chain (93 amino acids total), exon 1 of the cDNA SEQ ID NO: 29 making the difference
 between the 2 proteins (the last 3 exons show 100% homology). Thus, the protein of the invention
 represents a new isoform of human mitochondrial ATP synthase f subunit. It is interesting to note that
 the same splice variant is found in bovin, pig and mouse species.

15 The mitochondrial electron transport (or respiratory) chain is a series of enzyme complexes in
 the mitochondrial membrane that is responsible for the transport of electrons from NADH to oxygen
 and the coupling of this oxidation to the synthesis of ATP (oxidative phosphorylation). ATP then
 provides the

primary source of energy for driving a cell's many energy-requiring reactions. ATP synthase
 20 (F₀ F₁ ATPase) is the enzyme complex at the terminus of this chain and serves as a reversible coupling
 device that interconverts the energies of an electrochemical proton gradient across the mitochondrial
 membrane into either the synthesis or hydrolysis of ATP. This gradient is produced by other enzymes
 of the respiratory chain in the course of electron transport from NADH to oxygen. When the cell's
 energy demands are high, electron transport from NADH to oxygen generates an electrochemical
 25 gradient across the mitochondrial membrane. Proton translocation from the outer to the inner side of the
 membrane drives the synthesis of ATP. Under conditions of low energy requirements and when there is
 an excess of ATP present, this electrochemical gradient is reversed and ATP synthase hydrolyzes ATP.
 The energy of hydrolysis is used to pump protons out of the mitochondrial matrix. ATP synthase is,
 therefore, a dual complex, the F₀ portion of which is a transmembrane proton carrier or pump, and the
 30 F₁ portion of which is catalytic and synthesizes or hydrolyzes ATP. Mammalian ATP synthase
 complex consists of sixteen different polypeptides (Walker, J. E. and Collinson, T. R. (1994) *FEBS
 Lett.* 346: 39-43). Six of these polypeptides (subunits alpha, beta, gamma, delta, epsilon, and an ATPase
 inhibitor protein IF₁) comprise the globular catalytic F₁ ATPase portion of the complex, which lies
 outside of the mitochondrial membrane. The remaining ten polypeptides (subunits a, b, c, d, e, f, g, F₆,
 35 OSCP, and A6L) comprise the proton-translocating, membrane spanning F₀ portion of the complex.
 Like other members of the respiratory chain, all but two of the polypeptide subunits of ATP synthase

are nuclear gene products that are imported into the mitochondria. Enzyme complexes similar to mammalian ATP synthase are found in all cell types and in chloroplast and bacterial membranes. This universality indicates the central importance of this enzyme to ATP metabolism. Transcriptional regulation of these nuclear encoded genes appears to be the predominant means for controlling the biogenesis of ATP synthase. Multiple mitochondrial pathologies exist because of the essential role of mitochondrial oxidative phosphorylation in cellular energy production, in the generation of reactive oxygen species and in the initiation of apoptosis (Wallace, *Science*, 283:1482-1488, 1999). It is now clear that mitochondrial diseases encompass an assemblage of clinical problems commonly involving tissues that have high energy requirements such as heart, muscle and the renal and endocrine systems.

Over the past 11 years, a considerable body of evidence has accumulated implicating defects in the mitochondrial energy-generating pathway, oxidative phosphorylation, in a wide variety of degenerative diseases including myopathy and cardiomyopathy. Most classes of pathogenic mitochondrial DNA mutations affect the heart, in association with a variety of other clinical manifestations that can include skeletal muscle, the central nervous system (including eye), the endocrine system, and the renal system.

Nuclear mutations causing mitochondrial disorders have been described. They are often found in highly conserved subunits. Mitochondrial disorders with nuclear mutations include : myopathies (PEO, MNGIE, congenital muscular dystrophy, carnitine disorders), encephalopathies (Leigh, Infantile, Wilson's disease, Deafness-Dystonia syndrome), other systemic disorders and cardiomyopathies.

The discovery of a new ATP synthase subunit, and polynucleotides encoding it satisfy a need in the art by providing new compositions which are useful for the diagnosis, prevention, and treatment of cancer, myopathies, immune disorders, and neurological disorders.

It is believed that the protein of SEQ ID NO: 270 or part thereof plays a role in cellular respiration, preferably as a mitochondrial ATP synthase subunit. Preferred polypeptides of the invention are fragments of SEQ ID NO: 270 having any of the biological activity described herein.

An object of the present invention are compositions and methods of targeting heterologous compounds, either polypeptides or polynucleotides to mitochondria by recombinantly or chemically fusing a fragment of the protein of the invention to an heterologous polypeptide or polynucleotide. Preferred fragments are signal peptide, amphiphilic alpha helices and/or any other fragments of the protein of the invention, or part thereof, that may contain targeting signals for mitochondria including but not limited to matrix targeting signals as defined in Herrman and Neupert, *Curr. Opinion Microbiol.* 3:210-4 (2000); Bhagwat et al. *J. Biol. Chem.* 274:24014-22 (1999), Murphy *Trends Biotechnol.* 15:326-30 (1997); Glaser et al. *Plant Mol Biol* 38:311-38 (1998); Ciminale et al. *Oncogene* 18:4505-14 (1999). Such heterologous compounds may be used to modulate mitochondria's activities. For example, they may be used to induce and/or prevent mitochondrial-induced apoptosis or necrosis. In addition, heterologous polynucleotides may be used for mitochondrial gene therapy to replace a defective mitochondrial gene and/or to inhibit the deleterious expression of a mitochondrial gene.

The invention further relates to methods and compositions using the protein of the invention or part thereof to diagnose, prevent and/or treat several disorders in which mitochondrial respiratory electron transport chain is impaired, including but not limited to mitochondriocopathies, necrosis, aging, myopathies, cancer and neurodegenerative diseases such as Alzheimer's disease, Huntington's disease, Parkinson's disease, epilepsy, Down's syndrome, dementia, multiple sclerosis, and amyotrophic lateral sclerosis. For diagnostic purposes, the expression of the protein of the invention could be investigated using any of the Northern blotting, RT-PCR or immunoblotting methods described herein and compared to the expression in control individuals. For prevention and/or treatment purposes, the protein of the invention may be used to enhance electron transport and increase energy delivery using any of the gene therapy methods described herein or known to those skilled in the art.

In another embodiment, The invention further relates to methods and compositions using the protein of the invention or part thereof to diagnose, prevent and/or treat several disorders in which mitochondrial respiratory electron transport chain needs to be impaired, including but not limited to Sjogren's syndrome, Addison's disease, bronchitis, dermatomyositis, polymyositis, glomerulonephritis, diabetes mellitus, emphysema, Graves' disease, atrophic gastritis, lupus erythematosus, myasthenia gravis, multiple sclerosis, autoimmune thyroiditis, ulcerative colitis, anemia, pancreatitis, scleroderma, rheumatoid and osteoarthritis, asthma, allergic rhinitis, atopic dermatitis, dermatomyositis, polymyositis, and gout, using any techniques known to those skilled in the art including the antisense or triple helices strategies described herein.

Moreover, antibodies to the protein of the invention or part thereof may be used for detection of mitochondria organelles and/or mitochondrial membranes using any techniques known to those skilled in the art.

Protein of SEQ ID NO: 271 (internal designation 117-001-5-0-G3-CS)

The protein of SEQ ID NO: 271 is homologous to the family of lipopolysaccharide (LPS) binding proteins (LBPs). Several families of proteins have the ability to bind LPS including (a) the lipopolysaccharide-binding proteins (LBPs), and (b) the bactericidal permeability-increasing proteins (BPIs). Cholesteryl ester transfer protein (CETP), which is involved in the transfer of insoluble cholesteryl esters in reverse cholesterol transport, shares some homology to members of the LPS binding family of proteins.

Lipopolysaccharide (LPS), alternatively known as bacterial endotoxin, is a major component of the outer membrane of Gram-negative bacteria. It consists of serotype-specific O-side chain polysaccharides linked to a core oligosaccharide and Lipid A. LPS is a potent mediator of the inflammatory response and stimulates the expression of many pro-inflammatory and pro-coagulant compounds in monocytes, macrophages, and endothelial cells. While these responses are important in containing and eliminating localized infections, systemic exposure to LPS can lead to a number of

adverse effects. These include: (a) induction of an inflammatory cascade, (b) damage to the endothelium, (c) widespread coagulopathies, and (d) organ damage.

Systemic exposure to LPS can arise from direct infection by Gram-negative bacteria, leading to the complications of Gram-negative sepsis. Examples of diseases which are associated with Gram-negative bacterial infection or endotoxemia (including bacterial meningitis, neonatal sepsis, cystic fibrosis, inflammatory bowel disease, and liver cirrhosis), Gram-negative pneumonia, Gram-negative abdominal abscess, hemorrhagic shock, and disseminated intravascular coagulation. Subjects who are leukopenic or neutropenic, including subjects treated with chemotherapy or immunocompromised subjects, are particularly susceptible to bacterial infection and the subsequent effects of endotoxin exposure.

Gram-negative sepsis remains one of the primary causes of severe systemic inflammation in hospitalized and immunocompromised patients. Alternatively, changes in gut permeability by a variety of circumstances, including trauma, can lead to translocation of bacteria/LPS into the bloodstream. Bacteria translocated from the gut is thought to play a major role in post-surgical immunosuppression (Little et al., Surgery. 114: 87-91 (1993)) and hemorrhagic shock. Therefore, there is a great interest to characterize proteins involved in the biological response to LPS and to discover therapies that can counteract the effects of LPS in pathological situations.

LBP is a 60 kDa glycoprotein synthesized in the liver and present in normal human serum. LBP expression is upregulated in response to infectious, inflammatory, and toxic mediators. LBP expression has been induced in animals challenged with LPS, silver nitrate, turpentine, and *Corynebacterium parvum* (Geller et al., Surgery 128:22-28 (1993); Gallay et al., Infect. Immun. 61:378-383 (1993); Tobias et al., J. Exp. Med. 164:77-793 (1986)). LBP levels are correlated with exposure to LPS, and elevated levels (particularly persistent elevated levels) have been correlated with poor clinical outcomes in septic patients (U.S. Patent Nos. 5,484,705, and 5,804,367, hereby incorporated by reference in their entirety).

A portion of the LBP molecule (the N-terminal 1-197 aa) binds to the lipid A portion of the LPS molecule to form a high affinity LBP/LPS complex (Tobias, et al., J. Biol. Chem 264: 10867-10871 (1989)). The LBP/LPS complex potentiates the cellular response to LPS via an interaction with the monocytic differentiation antigen CD14 (Wright et al., Science. 249: 1431-1433 (1990); Lee et al., J. Exp. Med. 175:1697-1705 (1992)). LPS can be transferred from LBP to membrane-bound or soluble CD14. Activated CD14 can then interact with endothelial cells to elicit an inflammatory response. The C-terminal portion of LBP is required to transfer LPS to CD14 (U.S. Pat. No. 5,731,415; Theofan et al., J. Immunol. 152:3624-29 (1994); Han et al., J. Biol. Chem. 269:8172-75 (1994)). Evidence also suggests that LBP can neutralize LPS by an interaction with serum lipoproteins or through the internalization of an LBP/LPS/CD14 complex by neutrophils (Wurfel et al., J. Exp. Med. 180:1025-1035 (1994); Wurfel et al., J. Exp. Med. 181:1743-54 (1995); Gegner et al., J. Biol. Chem. 270:5320-5325 (1995)).

The subject invention provides the polypeptide of SEQ ID NO: 271 and polynucleotide sequences encoding the amino acid sequence of SEQ ID NO: 271. In a one embodiment, the polypeptides of SEQ ID NO: 271 are interchanged with the polypeptides encoded by the human cDNA of clone 181-20-3-0-B5-CS. Also included in the invention are biologically active fragments of the protein of SEQ ID NO: 271 and polynucleotide sequences encoding these biologically active fragments. In a preferred embodiment, biologically active fragments of SEQ ID NO: 271 are encoded by clone 181-20-3-0-B5-CS and comprise the first 181 amino acids encoded by clone 181-20-3-0-B5-CS. "Biologically active fragments" are defined as those peptide or polypeptide fragments of SEQ ID NO: 271 which have at least one of the biological functions of the full length protein (e.g., the ability to bind bacterial LPS).

The invention also provides variants of SEQ ID NO: 271. These variants have at least about 80%, more preferably at least about 90%, and most preferably at least about 95% amino acid sequence identity to the amino acid sequence of SEQ ID NO: 271. Variants according to the subject invention also have at least one functional or structural characteristic of SEQ ID NO: 271, such as the biological functions described above. The invention also provides biologically active fragments of the variant proteins. Unless otherwise indicated, the methods disclosed herein can be practiced utilizing the polypeptide of SEQ ID NO: 271 or variants thereof. Likewise, the methods of the subject invention can be practiced using biological fragments of the protein of SEQ ID NO: or variants of said biologically active fragments.

Because of the redundancy of the genetic code, a variety of different DNA sequences can encode SEQ ID NO: 271. It is well within the skill of a person trained in the art to create these alternative DNA sequences which encode proteins having the same, or essentially the same, amino acid sequence. These variant DNA sequences are, thus, within the scope of the subject invention. As used herein, reference to "essentially the same sequence" refers to sequences that have amino acid substitutions, deletions, additions, or insertions that do not materially affect biological activity. Fragments retaining one or more characteristic biological activity of SEQ ID NO: are also included in this definition.

"Recombinant nucleotide variants" are alternate polynucleotides which encode a particular protein. They can be synthesized, for example, by making use of the "redundancy" in the genetic code. Various codon substitutions, such as the silent changes which produce specific restriction sites or codon usage-specific mutations, can be introduced to optimize cloning into a plasmid or viral vector or expression in a particular prokaryotic or eukaryotic host system, respectively.

The protein of SEQ ID NO: 271, and variants thereof, can be used to produce antibodies according to methods well known in the art. The antibodies can be monoclonal or polyclonal. Antibodies can also be synthesized against fragments of SEQ ID NO: 271, as well as variants thereof, according to known methods. The subject invention also provides antibodies which

specifically bind to biologically active fragments of SEQ ID NO: 271 or biologically active fragments of SEQ ID NO: 271 variants.

The subject invention also provides for immunoassays which are used to screen for, monitor, or diagnose exposure to LPS. In one embodiment, diagnostic assays measure the level of LBP in patient plasma samples. LBP levels are known to rise in response to exposure to LPS, thus the measurement of the level of the protein of SEQ ID NO: 271 can provide an early indication of Gram-negative infection or of endotoxin exposure.

The subject invention provides methods of treating individuals infected with Gram negative bacteria comprising the administration of therapeutically-effective compositions comprising SEQ ID NO: 271. In one embodiment, the protein lacks the C-terminal portion (or portions of the C-terminal domain) necessary to transfer LPS to CD14. LPS can be scavenged by the excess N-terminal fragment and would be unable to induce an inflammatory response (see, e.g., U.S. Patent No. 5,731,415, hereby incorporated by reference in its entirety).

Another aspect of the subject invention provides methods of prophylaxis. The method treats individuals by administration of therapeutically-effective amounts of compositions comprising SEQ ID NO: 271. Instances where this aspect of the invention can be performed include, but are not limited to, conditions associated with increased translocation of gut bacteria and endotoxin, particularly prior to surgery. In addition, patients who are at risk for potential Gram-infection, including but not limited to patients undergoing chemotherapy, or patients who are immunocompromised (for example with AIDS) can benefit from such treatment. Such uses are described in U.S. Patent No. 5,990,082, hereby incorporated by reference in its entirety.

The N-terminal portion of LBP, which lacks the ability to induce an inflammatory response, can be fused to other proteins or fragments thereof (such as the bactericidal/permeability-increasing protein or BPI) which can increase the association of these molecules with LPS and aid in the clearance of endotoxin from patients who have been exposed to Gram negative bacteria. Such preparations can be used to treat and inhibit a number of Gram-negative infections, Gram positive, or fungal infections, as described in the following patents: WO 95/19179 A, WO 95/19180 A, WO 95/19372 A, and WO 96/34873 A, each of which is incorporated by reference in its entirety.

The subject invention also provides methods of removing endotoxin from recombinantly-produced proteins. In one embodiment, the recombinantly-produced proteins are obtained from Gram-negative bacteria. In a preferred embodiment, the bacteria are *E. coli*. In another embodiment, the protein of SEQ ID NO: 271, biologically active fragments thereof, variants, or derivatives thereof, are contacted with compositions comprising recombinantly-produced proteins. The contacting step can take place with SEQ ID NO: 271 immobilized on a substrate or with SEQ ID NO: 271 present in free solution.

In addition, protein of SEQ ID NO: 271, biologically active fragments, or derivatives thereof, can be used in diagnostic assays to measure the level of LPS in patient plasma samples. In

such an assay, serum samples would be bound to a solid matrix, such as a membrane, plastic, treated plastic, or other supports, and then cloned with the protein of SEQ ID NO: 271.

Visualization can be achieved by fusing protein of SEQ ID NO: to any number of enzymes followed by treatment with a chromogenic, fluorogenic, or luminescent substrate. Alternatively, the protein of SEQ ID NO: 271, biologically active fragments, variants, or derivatives thereof, can be linked to a fluorescent or luminescent protein or compound. The linkage can be chemical or made by recombinant techniques known to those skilled in the art. In addition, antibodies raised against the protein of SEQ ID NO: 271, biologically active fragments, variants, or derivatives thereof can be used to visualize the LPS/protein 271 complexes using immunoassays known to those skilled in the art.

Protein of SEQ ID NO:266 (internal designation 116-110-2-0-F4-CS)

The protein of SEQ ID NO:266, highly expressed in the testis, is encoded by cDNA of SEQ ID NO:25 and exhibits homology to the Ly-6 family of GPI-linked cell-surface glycoproteins composed of one or more copies of a conserved domain of about 100 amino-acid residues (PS00983; LY6_UPAR).

The protein of SEQ ID NO:266 shows significant structural similarities to mouse Ly-6 antigens, human CD59 and a herpes virus CD59 homolog. The protein of SEQ ID NO:266 displays one copy of the motif of the u-PAR/Ly-6 domain, with all ten extracellular cysteine residues conserved. The mature protein sequence contains a relatively high proportion of cysteine residues (10/105), which suggests that numerous disulfide bonds stabilize its tertiary structure. Furthermore, the 124 amino-acid long protein of SEQ ID NO:266 has a size very similar to that of many members of the Ly-6 family. In addition, the protein of the invention has a predicted signal peptide structure (positions from 1 to 19) and a C-terminal hydrophobic fragment (positions from 101 to 121) necessary for GPI-anchoring in a membrane. Thus, the protein of the invention has a clear evolutionary relationship with the Ly-6/uPAR family, particularly with Ly-6 subfamily.

The Ly-6/uPAR protein family members share one or several repeat units of the Ly-6/uPAR domain, which is defined by a distinct disulfide bonding pattern between 8 or 10 cysteine residues. This family can be divided into two subfamilies. One comprises GPI-anchored glycoprotein receptors with 10 cysteine residues. Another subfamily includes the secreted single-domain snake and frog cytotoxins, and differs significantly in that its members generally possess only eight cysteines and no GPI-anchoring signal sequence (Andermann K, et al. Protein Sci 8(4):810-819 (1999)). The Ly-6 family members are low molecular weight phosphatidyl inositol anchored glycoproteins with remarkable amino acid homology throughout a distinctive cysteine rich protein domain that is associated predominantly with O-linked carbohydrate. Their GPI links are necessary to anchor these cell surface proteins to the outside of the lipid bilayer membrane. The Ly-6 family includes human CD59, which protects from complement-mediated membrane damage, squid Sgpl

and Sgp2, urokinase plasminogen activator receptor, murine Sca-1 and Sca-2, and many other proteins. The general structure seen within the Ly-6 family resembles that of the receptor for a urokinase-type plasminogen activator and the alpha- neurotoxins from snake venoms (Fleming T J et al J Immunol 150:5379-5390 (1993); Ploug M and V Ellis FEBS Lett 349:163-168 (1994)).

5 The Ly-6 cell surface proteins are differentially expressed in several hematopoietic lineages that appear to function in signal transduction and cell activation predominantly on lymphoid cells in the mouse. Analyses using anti-Ly-6A/E monoclonal antibodies has also demonstrated in situ expression of Ly-6 molecules in brain tissue (staining primarily associated with vascular elements throughout the brain). These proteins do not appear to be expressed during embryonic or neonatal
10 stages of development (Cray C et al. Brain Res Mol Brain Res 8(1):9-15 (1990)).

 Ly-6 protein expression has been shown to be factor-dependent. For example, the expression of the Ly-6A/E, which normally occurs in hemopoietic stem cells, fibroblasts, and T and B lymphocytes, has been shown to be greatly induced by IFN- β in various tissues and cell lines. In addition, the Ly-6E Ag is associated with tyrosine kinases in T cells, and reduced expression of Ly-
15 6E in T cells impairs normal functional responses, as well as tyrosine kinase activity, in these cells. Further, the IFNs are important in the generation of memory CD8+ T cells, and it has been demonstrated that the expression of Ly-6C Ag is a strong marker for the memory phenotype (Mehran M. et al. Journal of Immunology 163: 811-819 (1999)). Like their murine counterparts, a human homologue of Ly-6 genes, the 9804 gene, is responsive to IFNs. The 9804 gene is also
20 inducible by retinoic acid during differentiation of acute promyelocytic leukemia cells. Further, cultured glial and neuronal cells express high levels of Ly-6A/E following incubation with cytokines, including rIFN-gamma. (Cray C et al. Brain Res Mol Brain Res 8(1):9-15 (1990)). Another member of the Ly-6 family, human protein RoBo-1, shows increased expression in response to two modulators of bone metabolism, estradiol and intermittent mechanical loading,
25 suggesting a role in bone homeostasis (Noel LS et al. J Biol Chem, Vol. 273(7): 3878-3883 (1998)). Such factor-dependence of expression makes Ly-6 proteins either candidates or targets for alloresponses and autoimmune disease. For example, the high level factor-induced expression of LY-6s has been associated with lupus nephritis (Blake P G et al. J Am Soc Nephrol 4:1140-1150 (1993)).

30 Murine Ly-6 molecules have interesting patterns of tissue expression during haematopoiesis, from multipotential stem cells to lineage committed precursor cells, and on specific leukocyte subpopulations in the peripheral lymphoid tissues. These patterns suggest an intimate association between the regulation of Ly-6 expression, and the development and homeostasis of the immune system (Gumley TP et al. Immunol Cell Biol 73(4):277-296 (1995)). Ly-6M messenger
35 RNA (mRNA) is easily detectable in hematopoietic tissue (bone marrow, spleen, thymus, peritoneal macrophages) as well as kidney and lung (Patterson JM et al. Blood 95(10):3125-3132 (2000)).

Normally, human blood cells are protected against autologous complement activation by membrane proteins that block the assembly of functional complement pores. One such protein is human Ly-6 CD59. Administration of CD59 prevents hemolytic disease or thrombosis. Further, the CD59 protein may prevent the complement-mediated lysis and activation of endothelial cells that leads to hyper acute rejection, and therefore may be administered during xenogeneic organ transplantation (Binette, J. P. and Binette, M. B., *Scanning Microsc.*, 7:1107-10 (1993)).

The surface receptor for urokinase plasminogen activator (uPAR) has been recognized in recent years as a key molecule in regulating plasminogen mediated extracellular proteolysis. Surface plasminogen activation controls the connections between cells, basement membrane and extracellular matrix, and therefore the capacity of cells to migrate and invade neighboring tissues (Roldan AL et al. *EMBO J* 9(2):467-474 (1990)). Certain factors of the PA system, such as u-PAR, have been detected in organs of the male reproductive tract in various species. The morphological study provide support for the involvement of the PA system in human male reproductive physiology (Gunnarsson M et al. *Mol Hum Reprod* 5(10):934-940 (1999)).

LY-6 proteins have been suggested to play important roles in disorders such as cancers, nephropathies, autoimmune diseases, hemolytic disease, thrombosis, Alzheimer's disease, etc. Several members of the murine Ly-6 supergene family are clearly involved in the progression of certain mouse tumors, as their expression level is higher in highly malignant cells than in tumor cells with a lower malignancy phenotype. Sorting by flow cytometry of tumor cells to subpopulations expressing either high or low levels of Ly-6E.1 yielded cells expressing a high or a low malignancy phenotype, respectively. Further, it was shown that LY-6 is highly expressed on non-lymphoid tumor cells originating from a variety of tissues in mice. Upregulation or high expression is correlated with a more malignant phenotype which results in higher efficiency of local tumor production (Katz et al *Int J Cancer* 59:684-91 (1994)).

Cells derived from angiogenic tumors express a higher tumorigenicity phenotype and a higher capacity to produce artificial pulmonary metastases than cells from the poorly angiogenic tumors. These cells also express significantly higher levels of the lymphocyte activation protein Ly-6E, so the angiogenic phenotype appears to be coregulated with Ly-6 (Sagi-Assif O et al. *Immunol Lett* 54(2-3):207-13 (1996)). Some LY-6 proteins also block secretion of interleukin II (IL-2) which is an approved anticancer agent and a key regulatory hormone in cell-mediated immunity (Fleming T J and T R Malek *J Immunol* 153:1955-62 (1994)). IL-2 stimulates the proliferation of both T and natural killer cells and activates NK cells which can directly lyse freshly isolated, solid tumor cells.

The high malignancy, high Ly-6E.1-expressing cells also expressed high levels of the receptor for urokinase plasminogen activator (uPAR), whereas low malignancy, low Ly-6E.1-expressing cells also expressed low levels of uPAR. Transfection studies have indicated that uPAR is causally involved in conferring a high malignancy phenotype upon tumor cells expressing high

levels of Ly-6E.1. E48, a human homologue of the murine ThB Ly-6 protein, is expressed on head and neck squamous carcinoma cells. In E48-stimulated cells, the binding of E48 to its microenvironmental ligand appears to transduce a signal that up-regulates the expression of the FX enzyme in these cells, leading to an increase in the levels of GDP-L-fucose (Rinat Eshel et al. J Biol Chem, Vol. 275(17):12833-12840 (2000)). A congenital disorder of leukocyte adhesion to vascular endothelium termed LADII is reflected in a generalized fucose deficiency and major defects in leukocyte trafficking and function. Ly-6 loss-variants of a murine tumor exhibit alterations in the incorporation of fucose and mannose into cellular glycoconjugates (Witz IP J. Cell. Biochem. Suppl. 34:61-66 (2000)).

10 It is believed that the protein of SEQ ID NO:266 is a novel member of the Ly-6 protein family, and is thus a specific cell-surface glycoprotein antigen involved in signal transduction and cell activation, proliferation and differentiation. Preferred polypeptides of the invention are polypeptides comprising the amino acids of SEQ ID NO:266 from position 1 to position 18 and from position 19 to position 124. Other preferred polypeptides of the invention are any fragments
15 of SEQ ID NO:266 having any of the biological activities described herein.

In one embodiment, this invention relates to methods and compositions using the protein of the invention or part thereof as a marker protein to selectively identify tissues, preferably testis. For example, the protein of the invention or part may be used to synthesize specific antibodies using any technique known to those skilled in the art. Such tissue-specific antibodies may then be used to
20 identify tissues of unknown origin, such as forensic samples, differentiated tumor tissue that has metastasized to foreign bodily sites, etc., or to differentiate different tissue types in a tissue cross-section using immunochemistry.

Another embodiment of the present invention relates to methods of using of the protein of the invention or part thereof and related compounds and derivatives to diagnose developmental and
25 malignant disorders in tissues including urogenital tissues and other tissues of the reproduction system of both sexes. For example, a biological sample is obtained from a patient with cancer or at risk of developing cancer, and the level of SEQ ID NO:25 polynucleotides or encoded polypeptides is detected within the cells of the sample. The detection of an elevated level of the SEQ ID NO:25 polynucleotides or encoded polypeptides in the sample relative to a control level indicates the
30 presence of malignant cells within the patient. The expression of the protein of the invention can be investigated using any of a number of methods, including, but not limited to, Northern blotting, RT-PCR or immunoblotting.

Another embodiment of the invention relates to compositions and methods using the protein of the invention or part thereof in recombinant protein form as pharmacological agents in the
35 treatment of developmental and malignant disorders in tissues including urogenital tissues and in other tissues of human reproduction system. Particularly, the protein of the invention or part thereof can be used in the treatment of disorders which are manifested by male sterility.

In another embodiment of the invention, antibodies which bind to the protein of the invention or part thereof are used in the treatment of tumors, e.g., human urogenital tumors, especially to enhance the secretion of interleukin II, which is an approved anticancer agent and key regulatory hormone in cell-mediated immunity. Such antibodies can be used alone or bound to a substance capable of ablating or killing cells as a therapy for urogenital disorders or cancers in which the protein of the invention is overexpressed.

The protein of the invention or part thereof may also be used in the treatment of diseases which can require transplantation, including various forms of cancers such as genitourinary cancers, carcinomas, sarcomas, atherosclerosis, angiogenesis, and benign tumors. As mentioned above, Lysine 6 family includes several proteins which are similar to the protein of the invention and which are capable of protecting cells from complement-mediated membrane damage. Therefore, in another embodiment of the invention, recombinant proteins encoded by SEQ ID NO:25 or fragments thereof are administered during xenogeneic tissue transplantation to prevent complement-mediated lysis and to block activation of endothelial cells, which normally leads to hyper-acute rejection.

In addition, prevention of complement-mediated lysis may be particularly important in human and animal reproductive therapy, where functional survival of the germ cells during in vitro handling is crucial. Storage of sperm is of widespread importance in commercial animal breeding programs, human sperm donor programs, and in the treatment of certain disease states. For example, sperm samples may be frozen for men who have been diagnosed with cancer or other diseases that may eventually interfere with sperm production, as well as for assisted reproduction purposes where sperm may be stored for use at other locations or times. The procedures utilized in such cases include: washing a sperm sample to separate out the sperm-rich fraction from non-sperm components of a sample such as seminal plasma or debris; further isolating the healthy, motile sperm from dead sperm or from white blood cells in an ejaculate; freezing or refrigerating of sperm for use at a later date or for shipping to females at differing locations; extending or diluting sperm for culture in diagnostic testing or for use in therapeutic interventions such as in vitro fertilization or intracytoplasmic sperm injection (Cohen et al. 12 : 994-1001 (1997)). Once sperm have been washed or isolated, they are then extended (or diluted) in culture or holding media for a variety of uses (sperm analysis, diagnostic tests, assisted reproduction). Each of these uses for extended or diluted sperm requires a somewhat different formulation of basal medium (see, for review, US Patent No. 6,140,121 Ellington et al. Oct. 2000); however, in all cases sperm survival is suboptimal outside of the female reproductive tract. Novel additional components of a dilution or storage medium which could improve the functional preservation of sperm would be useful. Therefore, in another preferred embodiment of this invention, purified recombinant proteins encoded by SEQ ID NO:25 or fragments thereof can be added as components of pharmacological media designed to protect spermatozoa. The methods used to compose such preservation media are generally known by those skilled in the art (for ex., Oliver S.A. et al. US patent 5,897,987 Apr. 1999; Cohen J. et al.,

supra). Inversely, in yet another embodiment of this invention, ligands, inhibitors, neutralizing antibodies or other biological agents which recognize the protein of the invention and which bind it and which block it can be used as components of pharmacological formulations designed for male contraception purposes.

5 In still another embodiment of this invention, chimeric ligands or derivatives which recognize the protein of the invention or part thereof and which could be internalized into cell can be used to design a system of drug delivery finely targeted toward urogenital and other tissues which express the protein of SEQ ID NO:266. For example, such recognizing molecules can be incorporated into the membranes of liposomes to allow the specific delivery of the liposomes to
10 cells expressing the protein of SEQ ID NO:266. Methods of designing such drug delivery systems are known by those skilled in the art (Smith H.J. Introduction to the principles of drug design and action, 3rd ed. (1998)).

Proteins SEQ ID NOs:417, 413, 418 (internal designations 188-45-1-0-D3-CS, 188-26-4-0-F5-CS, and 188-5-1-0-H6-CS)

15 The proteins of SEQ ID NOs:417, 413, and 418, encoded by the cDNAs of SEQ ID NOs: 176, 172, and 177, are expressed in the brain and exhibit strong homology with proteins with redox activity (see, e.g. Genbank accession numbers AK001293 and AF029689, and Genesep accession number: Y59180).

The protein of SEQ ID No:418 (320 amino acids) is a variant of AK001293 (322 amino
20 acids). AK001293 has six extra nucleotides, within the same ORF, as SEQ ID No:418, producing a longer protein. SEQ ID NO:418 exhibits the Pfam Zinc-binding dehydrogenase (adh zinc) signature from positions 16 to 313. SEQ ID NO:418 presents all the conserved residues of the motif except for a histidine that is thought to be a zinc-ligand. This lack of zinc-ligand residues is a feature of the quinone oxidoreductases (QOR), a subfamily of zinc-binding dehydrogenases.

25 SEQ ID NO:413 (191 amino acids) shares the first 172 amino acids with SEQ ID NO:418. The deletion of one nucleotide at position 583 in the SEQ ID NO:413 cDNA sequence (corresponding to amino acid 173), however, creates a change of ORF compared to SEQ ID NO:418 and AK001293.

SEQ ID NO:417 is a short protein (20 amino acids) whose sequence corresponds to the N-
30 terminal end of the other proteins of the invention. The presence of a T (instead of a G in public sequences and SEQ ID NOs:413 and 418) at position 128 on the cDNA creates a STOP codon, creating a shorter protein.

SEQ ID NOs:417, 413 and 418 are similar to the QORs, a family of zinc-binding
dehydrogenases. QORs are cytoplasmic redox-regulated flavoenzymes that catalyze the one or two-
35 electron reduction of quinones. QORs bind NADP and are inhibited by dicoumarol.

The activity of QORs protects cells against toxicity, mutagenicity, and cancer due to exposure to environmental and synthetic quinones and their precursors. Thus, QORs play a central role in monitoring cellular redox state and act to protect against oxidative stress induced by a variety of metabolic situations (Raina A.K. et al. (1999) Redox Rep. 4:23-7). The oxidoreductase activity also permits the activation of bioreductive anticancer drugs (Begleiter A. et al. (1996) Br. J. Cancer Suppl. 27:S9-14).

The metabolism of quinones involves enzymatic reduction of the quinone by one or two electrons. In the activation of quinone-containing antitumor agents, this reduction results in the formation of the semiquinone or the hydroquinone of the anticancer drug. The consequence of these enzymatic reductions is that the semiquinone yields its extra electron to oxygen with the formation of superoxide radical anion and the original quinone. This reduction by a reductase followed by oxidation by molecular oxygen (dioxxygen) is known as redox-cycling and continues until the system becomes anaerobic. In the case of a two-electron reduction, the hydroquinone could become stable, and as such, be excreted by the organism in a detoxification pathway.

The cellular antioxidant response is mediated by a battery of detoxifying/defensive proteins. The promoters of genes that encode these proteins contain a common cis-element termed the antioxidant response element (ARE). Many transcription factors, including Nrf, Jun, Fos, Fra, Maf, YABP, ARE-BP1, Ah (aromatic hydrocarbon) receptor, and estrogen receptor bind to the ARE from various genes. Among these factors, Nrf-Jun heterodimers positively regulate ARE-mediated expression and induction of genes in response to antioxidants and xenobiotics (reviewed in Dhakshinamoorthy S. et al. (2000) Curr. Top Cell Regul. 36:201-16). On the other hand, c-Fos represses ARE-mediated gene expression (Venugopal, R., and Jaiswal, A.K. (1996) Proc. Natl. Acad. Sci. USA 93, 14960-5).

Elevated levels of QOR activity have been reported in several kinds of tumors such as liver, colon, lung and breast (Belinsky M., Jaiswal A.K., (1993) Cancer Metastasis Rev 12:103-17). Bioreactive antitumor agents are an important class of anticancer drugs that require activation by reduction. For this reason, QORs are a potential target on which to base the development of new antitumor compounds. Certain QORs have already been implicated in the metabolism, activation and mechanism of cytotoxicity of some anticancer drugs such as mitomycin C, indoloquinone E09 (Ross D. et al. (1994) Oncol. Res. 6:493-500), CB 1954 (Knox R.J. et al. (2000) Cancer Res. 60:4179-86) or antiestrogens in breast cancer (Montano M.M., Katzenellenbogen B.S. (1997) PNAS 94:2581-6).

In addition, some of the proteins of the QOR family are thought to play a role in the prevention of apoptosis following oxidative stress. The tumor suppressor gene p53 has been directly implicated in the induction of apoptosis in dividing cells and in hippocampal pyramidal neurons (Jordan J. et al. (1997) J. Neurosci 17:1397-405) and a QOR gene has been described as a p53-regulated gene (Kostic C, Shaw P.H. (2000) Oncogene 19:3978-87).

It is believed that the proteins of SEQ ID NOs:417, 413 and 418 have a redox activity, most likely as QORs. Thus, they are expected to act as an endogenous antioxidant against oxidative stress and may be able to use NADP as cofactor. The proteins of the invention may be used to deactivate toxins and to activate bioreductive anticancer drugs. In addition, they may prevent apoptosis following oxidative stress and be regulated by p53. Because proteins SEQ ID NOs:417 and 413 do not contain the Pfam Zinc-binding dehydrogenase (adh zinc) signature, in contrast to SEQ ID NO:418, they may act as a competitive inhibitor, i.e. a dominant negative form, of the functional protein.

The oxidoreductase activity of the proteins of the invention may be assayed using any technique known to those skilled in the art. For example, the measurement of the rate of oxidation of NADPH and oxygen consumption, and the detection of the semiquinone and reactive oxygen species, may be performed as described by Gutierrez P.L. (Gutierrez P.L. (2000) *Front. Biosci.* 5:D629-38), or by any other method skilled in the art. The enzymatic activity of the proteins of the invention in different affected and control tissues may be assayed by histochemical staining. To confirm the role of the proteins of the invention in the cellular antioxidant response, in vitro and in vivo assays may be performed. Transcription levels of the genes coding for the proteins of the invention may be measured using standard techniques after exposure to quinones or derived compounds as beta-naphthoflavone (beta-NF), as described by Belinsky M. and Jaiswal A.K. (supra), as well as in response to transcription factors such as Nrf, Jun and c-Fos, or in the presence of p53.

In one embodiment of the present invention, the present protein can be used to detect specific cell types in vitro or in vivo. For example, as the present proteins are overexpressed in the brain, reagents capable of specifically recognizing the present protein can be used as markers for brain cells. Brain-specific markers have a number of uses, including for the identification of specific tissues for histological analyses, as well as to detect the origin of tumor cells. In addition, as the expression of the present protein is likely induced by transcription factors such as Nrf, Jun, Fos, Fra, Maf, YABP, ARE-BP1, Ah (aromatic hydrocarbon) receptor, and estrogen receptor, as well as by p53, reagents specific for detecting the present protein can also be used as a marker for the activity of any of these proteins in vitro or in vivo. In view of the association between many of these proteins and diseases such as cancer, the ability to detect the presence or absence of the proteins provides powerful tools for disease diagnosis and screening. For any of these applications, the expression of the present protein can be detected using any standard method, including Northern blots, western blots, in situ hybridization, PCR, etc.

In another embodiment, the proteins of the invention can serve as markers for cellular oxidative stress in vivo and in vitro. As such, the proteins of the invention or part thereof may be useful in the diagnosis of disorders in which oxidative stress is implicated, including a large variety of types of cancer as well as neurodegenerative disorders such as Alzheimer's disease (AD), amyotrophic lateral sclerosis (ALS) or Parkinson disease (PD). For diagnostic purposes, the

expression of the protein of the invention may be investigated using, e.g. Northern blotting, RT-PCR or immunoblotting methods and compared to the expression in control individuals. An increased levels of the proteins of the invention in patients compared with controls indicates a major shift in redox balance and, thus, indicates the presence of the disease or of a susceptibility for the disease.

The invention further relates to methods and compositions using the proteins of the invention or part thereof to prevent and/or treat disorders in which oxidative stress is implicated, including those mentioned above. For these purposes the proteins themselves, or polynucleotides encoding the proteins, or an activator of protein expression may be administrated to patients, or to disease-free individuals in case of increased susceptibility to one of these disorders.

In another embodiment, the protein of the invention or part thereof is used to prevent cells from undergoing apoptosis. They may thus be useful in the diagnosis, treatment and/or prevention of disorders and processes in which apoptosis is deleterious, including but not limited to immune deficiency syndromes (including AIDS), type I diabetes, pathogenic infections, cardiovascular and neurological injury, alopecia, aging, degenerative diseases including AD and PD, dystonia, Leber's hereditary optic neuropathy and schizophrenia. For all such diagnostic purposes, the expression of the proteins of the invention can be investigated using any of the Northern blotting, RT-PCR or immunoblotting methods described herein and compared to the expression in control individuals.

The invention relates to methods and compositions using the proteins of the invention or part thereof as detoxifying enzymes against quinones. There are a variety of quinones with a toxic effect in cells (e.g. quinones derived from the oxidation of phenolic metabolites of benzene, DA-quinones, or menadione). Thus, the proteins of the invention or part thereof may be protective against the hematotoxic and carcinogenic effects of benzene, as well as against benzene-caused diseases such as cancer, aplastic anemia and pancytopenia. Moreover, they may detoxify DA-quinones in the brain, thereby providing neuroprotection in Parkinson's Disease. In still another embodiment, the proteins of the invention or part thereof may protect cells against menadione-induced oxidative stress, with known effects on myocardial cells (Floreani M. et al (2000) Biochem Pharmacol. 60:601-5). For prevention and/or treatment purposes the proteins themselves, or polynucleotides encoding the proteins, or an activator of protein expression may be administrated.

In another embodiment, the present proteins may be a target of chemotherapy specific to different kinds of cancer, to ensure a favorable response to anticancer drugs. Specifically, proteins of the invention or part thereof may be used as an activator of cytotoxic prodrugs of quinone family. Accordingly, the protein of the invention or part thereof may be administered to a patient in conjunction with a bioreductive anticancer agent in order to activate the drug. This co-administration may be by simultaneous administration, such as a mixture of the oxidoreductase and the drug, or by separate simultaneous or sequential administration. Cancer-specific antitumor agents based on QOR substrates may be designed as described by Xing J. et al. (Xing J. et al. (2000)

Med. Chem. 43:457-66) and assayed as described in Li B. et al. (Li B et al. (1999) Chem. Res. Toxicol. 12:1042-9). Alternatively, as the present proteins may be overexpressed in tumor cells, such methods may be performed by simply detecting the level of the present protein in tumor cells, and administering the prodrug specifically to those patients found to have elevated levels of the
 5 protein in their tumor cells.

Proteins of SEQ ID NOs: 415, 310, 317 (internal designation 188-29-2-0-H1-CS, 188-18-4-0-A9-CS, 188-9-2-0-E1-CS)

Mammalian inositol hexakiphosphate kinase 2 (IP6K2), an enzyme of the inositol phosphate pathway, has been cloned and described by two independent groups [Saiardi, A.;
 10 Erdument-Bromage, H.; Snowman, A. M.; Tempst, P.; and Snyder, S. H., (1999) Current Biology 9, 1323-1326, and Katai, K.; Miyamoto, K-I.; Kishida, S.; Segawa, H.; Nii, T.; Tanaka, H.; Tani, Y.; Arai, H.; Tatsumi, S.; Morita, K.; Taketani, Y.; and Takeda, E. (1999) Biochem. J. 343, 705-712]. Newly identified consensus sequences of inositol-polyphosphate kinases are represented by [LV]-[LA]-[DE]-X(3-8)-P-X-[VAI]-[ML]-D-X-K-[ML]G [Saiardi, A.; Erdument-Bromage, H.;
 15 Snowman, A. M.; Tempst, P.; and Snyder, S. H. (1999) Current Biology 9, 1323-1326]. IP6K2 catalyzes the transfer of phosphate groups from InsP6 or Ins(1,3,4,5,6)P5 (the substrate), to another protein or small molecule, such as a nucleoside di-phosphate.

The subject invention provides the polypeptides of SEQ ID NOs:415, 310, and 317, encoded by the cDNAs of SEQ ID NOs:174, 69, and 76, respectively. The invention also provides
 20 biologically active fragments of SEQ ID NOs:415, 310, and 317. In one embodiment, the polypeptides of SEQ ID NOs:415, 310, and 317 are interchanged with the corresponding polypeptides encoded by the human cDNA of clone 188-29-2-0-H1-CS, 188-18-4-0-A9-CS, or 188-9-2-0-E1-CS. "Biologically active fragments" are defined as those peptide or polypeptide fragments having at least one of the biological functions of the full length protein (e.g., kinase
 25 activity). Compositions of the protein/polypeptide of SEQ ID NOs:415, 310, or 317, or biologically active fragments thereof, are also provided by the subject invention. These compositions may be made according to methods well known in the art.

The invention also provides variants of the protein of SEQ ID NOs:415, 310, or 317. These variants have at least about 80%, more preferably at least about 90%, and most preferably at least
 30 about 95% amino acid sequence identity to the amino acid sequences encoded by SEQ ID NOs:415, 310, and 317. Variants according to the subject invention also have at least one functional or structural characteristic of the protein of SEQ ID NOs:415, 310, or 317. The invention also provides biologically active fragments of the variant proteins. Compositions of variants, or biologically active fragments thereof, are also provided by the subject invention. These
 35 compositions may be made according to methods well known in the art. Unless otherwise indicated, the methods disclosed herein can be practiced utilizing the protein encoded by SEQ ID

NO:415, 310, or 317, biologically active fragments of SEQ ID NO:415, 310, or 317, variants of SEQ ID NO:415, 310, or 317, and biologically active fragments of the variants.

Because of the redundancy of the genetic code, a variety of different DNA sequences can encode the amino acid sequence of SEQ ID NO:415, 310, or 317. In a preferred embodiment, SEQ ID NO:415, 310, or 317 is encoded by clone 188-29-2-0-H1-CS, 188-18-4-0-A9-CS, or 188-9-2-0-E1-CS, or by the cDNAs of SEQ ID NO:174, 69, or 76. It is well within the skill of a person trained in the art to create these alternative DNA sequences which encode proteins having the same, or essentially the same, amino acid sequence. These variant DNA sequences are, thus, within the scope of the subject invention. As used herein, reference to "essentially the same" sequence refers to sequences that have amino acid substitutions, deletions, additions, or insertions that do not materially affect biological activity. Fragments retaining one or more characteristic biological activity of the protein encoded by SEQ ID NO:415, 310, or 317 are also included in this definition.

In one aspect of the subject invention, SEQ ID NO:415, 310, or 317, and variants thereof, can be used to generate polyclonal or monoclonal antibodies. Both biologically active and immunogenic fragments of SEQ ID NO:415, 310, or 317, or variant proteins, can be used to produce antibodies. Polyclonal and/or monoclonal antibodies can be made according to methods well known to the skilled artisan. Antibodies produced in accordance with the subject invention can be used in a variety of detection assays known to those skilled in the art. The antibodies may be used to agonize or antagonize the biological activity of the protein of SEQ ID NO:415, 310, or 317.

The protein of SEQ ID NO:415, 310, or 317 can be used for the synthesis of nucleoside triphosphate (NTP) compounds. In one embodiment, the NTP compound produced is ATP, GTP, CTP, or TTP. In this aspect of the subject invention, SEQ ID NO:415, 310, or 317 removes a phosphate from InsP6 or Ins(1,3,4,5,6)P5 and transfers it to a nucleoside diphosphate (e.g., ADP, CTP, GDP, or TDP) to create a NTP. The conditions and methods for the synthesis of NTP compounds, such as ATP, are well known to the skilled artisan. Thus, the protein of SEQ ID NO:415, 310, or 317 has industrially useful function for the synthesis of commercially valuable products.

The subject invention also provides methods of determining the relative amounts of InsP6 or Ins(1,3,4,5,6)P5 in the cell by a kinase assay. In this aspect of the invention, SEQ ID NO:415, 310, or 317 can be used to transfer phosphate groups from InsP6 or Ins(1,3,4,5,6)P5 to acceptor substrates according to well-known kinase activity assays.

Protein of SEQ ID NO:294 (internal designation 181-16-2-0-A7-CS)

The protein of SEQ ID NO:294 is encoded by the cDNA of SEQ ID NO:53. It will be appreciated that all characteristics and uses of the polypeptide of SEQ ID NO:294 described throughout the present application also pertain to the polypeptide encoded by the human cDNA of clone 181-16-2-0-A7-CS. In addition, it will be appreciated that all characteristics and uses of the

nucleic acid of SEQ ID NO:53 described throughout the present application also pertain to the human cDNA of clone 181-16-2-0-A7-CS.

This gene was isolated from fetal liver and expression has also been detected in fetal kidney, placenta, liver, brain, hypertrophic prostate, salivary gland and testis. Data from PCT application WO 98/23435 indicate expression is primarily in bone marrow cell lines, and to a lesser extent, in human endometrial stromal cells, human adult small intestine and human pancreas tumor. PCT application WO 99/14484 reports the fraction of expression in the gastrointestinal system (0.227), reproductive system (0.193), and hematopoietic/immune system (0.168). Finally, this protein is 55% identical and 76% similar to CGI-128 protein, which was isolated from CD34+ cells and is also found in cell lines from the hematopoietic lineage including, HL6 (granulocytic), Jurkat (T-lymphocytic), K562 (erythro-megakaryocytic), and U937 (monocytic).

Supernatant harvested from cells expressing the product of this gene has been shown to increase the permeability of the plasma membrane of renal mesangial cells to calcium. Thus, it is believed that the product of this gene is involved in activating a signal transduction pathway when it binds a receptor on the surface of the plasma membrane of both mesangial cells and other cell types, in addition to other cell-lines or tissue cell types. Thus, polynucleotides and polypeptides have uses, which include, but are not limited to, activating mesangial cells by contacting said cells with a full length polypeptide or a polypeptide fragment which demonstrates this biological activity. Further, the polynucleotides and polypeptides can be used in the methods described in WO9915652, incorporated in its entirety. Binding of a ligand to a receptor is known to alter intracellular levels of small molecules, such as calcium, potassium and sodium, as well as alter pH and membrane potential. Alterations in small molecule concentration can be measured to identify supernatants, which bind to receptors of a particular cell. In addition, when tested against fibroblast cell lines, supernatants removed from cells containing this gene activated the EGR1 (early growth response gene 1) promoter element. Thus, it is likely that this gene activates fibroblast cells through the EGR1 signal transduction pathway. EGR1 is a separate signal transduction pathway from Jak-STAT, genes containing the EGR1 promoter are induced in various tissues and cell types upon activation, leading the cells to undergo differentiation and proliferation (PCT application WO 98/23435)

Polynucleotide comprising sequences encoding the signal peptide of the protein, e.g. VLWLSGLSEPGAA/RQ, can be used in construction of secretion vectors. These vectors would then facilitate the secretion of fusion proteins into the media of cells that have been transfected with the construct of interest. Antibodies which specifically bind the signal peptide could be used to purify the fusion protein from the media if desired.

Polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, haemopoietic and gastrointestinal tract

disorders and stromatosis, in addition to endothelial, mucosal, or epithelial cell disorders. Similarly, polypeptides and antibodies directed to these polypeptides, are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and digestive systems, expression of this gene at
5 significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g. hemaopoietic, immune, reproductive, gastrointestinal, endocrine, and cancerous and wounded tissues) or bodily fluids (e.g. lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an
10 individual not having the disorder.

The tissue distributioin in bone marrow cells, fetal liver and fetal kidney, combined with the detected calcium flux and EGR1 biological activity, indicates that polynucleotides and polypeptides corresponding to this gene are useful for immune and gastrointestinal tract disorders, and stromatosis, particularly tumors and proliferative disorders. More specifically, polynucleotides and
15 polypeptides corresponding to this gene are useful for the treatment and diagnosis of hematopoietic related disorders such as anemia, pancytopenia, leukopenia, thrombocytopenia or leukemia since stromal cells are important in the production of cells of hematopoietic lineages. The polypeptides and polynucleotides of the invention can be used to enhance hematopoiesis as described in WO9831385, incorporated in its entirety. The uses include bone marrow cell ex vivo culture, bone
20 marrow transplantation, bone marrow reconstitution, radiotherapy or chemotherapy of neoplasia. The gene product may also be involved in lymphopoiesis, therefore, it can be used in immune disorders such as infection, inflammation, allergy, immunodeficiency etc. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Protein
25 as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Additionally, since the gene product of 181-16-2-0-A7-CS has been shown to activate the EGR1 promoter element, it likely activates EGR1 signaling activity in fibroblasts. Recent data shows that activation of EGR1 plays a role in wound repair. The cellular transcription factor early
30 growth response factor 1 (Egr1) is expressed minutes after acute injury and serves to stimulate the production of a class of growth factors whose role is to promote tissue repair. Egr-1 expression at the site of dermal wounding in rodents promotes angiogenesis in vitro and in vivo, increases collagen production, and accelerates wound closure. These results show that Egr-1 gene therapy accelerates the normal healing process (Human Gene Ther 2000, vol 11(15):2143-58). Thus, an
35 activator of EGR1 signaling, specifically the gene products of 181-16-2-0-A7-CS (polypeptides and polynucleotides), would be useful in the wound healing process using the methods described in WO9941282 and WO9932135, incorporated by reference in their entirety.

Protein of SEQ ID NO:305 (internal designation 187-37-0-0-c10-CS)

The protein of SEQ ID NO:305, encoded by the cDNA of SEQ ID NO:64, is highly expressed in the prostate and brain. The protein of the invention is strongly homologous to the D9 protein, found in both humans (GNP accession number: U95006 and U95007) and in mice (GNP
5 accession number: U95003, U95004, and U95005). D9 is a myeloid precursor protein transcript regulated by the retinoic acid receptor α , hereafter referred to as RAR- α (Scott et al. Blood 1996; 88: 2517-30).

Retinoic acid is the active metabolite of vitamin A, which contributes to a wide range of biological processes such as cellular differentiation, embryogenesis, and tumor suppression. More
10 specifically, retinoic acid stimulates myeloid precursor differentiation into mature granulocytes. For instance, *in vitro* treatment of acute promyelocytic leukemia blast cells with retinoic acid induces their differentiation (Miyauchi et al. Leuk Lymphoma 1999;33:267-80). In addition, treatment with retinoic acid can induce disease remission in patients affected with promyelocytic leukemia by causing granulocyte precursor differentiation (Slack et al. Ann Hematol 2000;79:227-
15 38).

The diverse range of responses to retinoic acid are mediated by three receptor subtypes: RAR- α , RAR- β , and RAR- γ . RAR- α has been identified as being important for bone marrow maturation of granulocytes (Tsai et al. Genes Dev 1992;6:2258-69). In addition, RAR- α is almost invariably involved in acute promyelocytic leukemia cells by a reciprocal translocation between the
20 long arms of chromosomes 15 and 17 (Alcalay et al., Proc Natl Acad Sci USA 1991;88:1977-81). This type of leukemia is mainly characterized by a predominance of malignant promyelocytes, and severe hemorrhagic manifestations resulting from activation of the coagulation cascade and the fibrinolytic system (Tallman et al. Semin Thromb Hemost 1999;25:209-15). Reciprocal chromosomal translocation leads to the production of a fusion protein that inhibits differentiation
25 and promotes survival of myeloid precursor cells (Grignani et al. Cell 1993;74, 423-431). Transient transfection of a vector containing RAR- α in a promyelocyte cell line causes upregulation in an early manner of several genes, including D9, which is strongly related to protein of SEQ ID NO:305 (Scott et al. Blood 1996; 88: 2517-30). Thus, it is believed that the protein of SEQ ID NO:305 is a myeloid-related protein whose expression is induced by the activation of retinoic acid receptors,
30 including RAR- α .

In a preferred embodiment, the protein of the invention or part thereof may be used to assay the activity of RAR- α protein or retinoic acid in a biological sample. Specifically, as the expression of the protein is believed to be under the direct control of retinoic acid receptors, the level of the protein of the invention, or of the mRNA encoding the protein, can serve as a sensitive and
35 immediate marker for the effects of retinoic acid upon a cell. An ability to detect retinoic acid receptor activation in cells using the present protein has numerous uses. For instance, the protein of the invention or part thereof can be used to monitor the effects of retinoic acid on cells of a patient

undergoing retinoic acid treatment for promyelocytic leukemia (Slack et al. Ann Hematol 2000;79:227-38). As retinoic acid treatment is associated with frequent retarded dose-dependant side effects, it is believed that an assay based on protein of SEQ ID NO:305 could be used to adjust the dose of retinoic acid administered in patients affected with promyelocytic leukemia, in order to
5 predict and avoid such adverse side-effects (Slack et al. Ann Hematol 2000;79:227-38).

In another embodiment, the present polypeptides and polynucleotides can be used to identify myeloid precursors, as well as brain and prostate tissues. The ability to specifically visualize myeloid precursor cells, as well as brain and prostate tissues (and cells derived from the tissues), is useful for any of a number of applications, including to determine the origin or identity
10 of, e.g. cancerous cells, as well as to facilitate the identification of particular cells and tissues for, e.g. the evaluation of histological slides. In addition, such assays can be used to examine the extent of differentiation in myeloid precursor cells.

The present invention further relates to *in vitro* assays and diagnostic kits based on the protein of the present invention or part thereof. Such assays may be used for diagnosis of disorders
15 where the protein activity is abnormally downregulated, such as cancer, and hematological disorders including leukemia. As the protein of SEQ ID NO:305, RAR- α , and acute promyelocytic leukemia are all related, variation in the measured level of the present protein of the invention or part thereof can be used as a diagnostic or screening test for acute promyelocytic leukemia, e.g. using a biological sample such as serum or plasma. Further, an assay that can detect an abnormal
20 level of the protein of the invention or part thereof can be used to detect residual disease in acute promyelocytic leukemia. Such an assay may be used to aid therapeutic decisions in this disorder, e.g. more or less aggressive treatments, the duration of treatment, etc.

In another embodiment, various methods can be used to modulate activity and/or expression of the protein of SEQ ID NO:305, e.g. for the treatment, attenuation and/or prevention of various
25 disorders. In one embodiment, any of a number of reagents, e.g. polynucleotides encoding the protein of SEQ ID NO:305 or a fragment thereof, the protein of SEQ ID NO:305 itself, or a compound that increased the expression or activity of the protein of SEQ ID NO:305, can be administered to a patient suffering from, or at risk of developing, various disorders including cancer, and hematological diseases such as leukemia, and neutropenia. For instance, but not limited
30 to it, proteins or other capable of enhancing the expression or activity of the protein of SEQ ID NO:305 can be administered to treat patients affected with acute promyelocytic leukemia, in order to induce differentiation of the affected cells into mature granulocytes (Slack et al. Ann Hematol 2000;79:227-38). In still another preferred embodiment, proteins or other compounds capable of increasing the expression or activity of the protein of the invention can be used to treat, prevent
35 and/or attenuate neutropenia or agranulocytosis patients, in order to induce *in vivo* differentiation of myeloid precursors into mature granulocytes. In still another preferred embodiment, proteins or other compounds capable of increasing the expression or activity of the protein of SEQ ID NO:305

can be used to treat coagulopathic diseases, such as thrombosis or hemorrhagic manifestations. For instance, they can be used to treat disseminated intravascular coagulation, a severe hemorrhagic syndrome. This embodiment is supported by the fact that acute promyelocytic leukemia is frequently associated with disseminated intravascular coagulation (Tallman et al. Semin Thromb Hemost 1999;25:209-15), and disseminated intravascular coagulation is efficiently corrected with retinoic acid (Dombret et al. Leukemia 1993;7:2-9).

In addition, modulation of the expression or activity of the protein of the invention can be used to modulate differentiation of cells, e.g. promyelocytic leukemia. In one such embodiment, the protein of the invention is inhibited, e.g. using antisense molecules, antibodies, or small molecule inhibitors of the expression or activity of the protein, in order to maintain the undifferentiated state of cells grown in vitro. Alternatively, agents that increase the expression or activity of the protein in cells can be used to induce cellular differentiation, e.g. in the preparation of specific cell types in vitro for particular therapeutic applications.

Protein of SEQ ID NO:248 (internal designation 105-035-2-0-C6-CS) and SEQ ID NO:313 (internal designation 188-28-4-0-D4-CS)

The proteins of SEQ ID NO:248, encoded by the cDNA of SEQ ID NO:7, and SEQ ID NO:313, encoded by the cDNA of SEQ ID NO:72, are highly expressed in brain, liver, pancreas, and testis. The proteins of the invention are nuclear proteins (Miller et al. J Biol Chem 2000;275:32052-6) that display a membrane-spanning segment from amino acids 58 to 78. These proteins are homologous to the human RNA polymerase II elongation factor ELL3 (EMBL accession number AF276512 ; Miller et al. J Biol Chem. 2000; 275:32052-6). In addition, the proteins of SEQ ID NO:248 and SEQ ID NO:313 share sequence homology with two other members of the polymerase II elongation factor family: ELL, and ELL2. The protein of SEQ ID NO:313 is similar to the N-terminal sequence the protein of SEQ ID NO:248, but differs after residue 240 because of a frameshift that produces a premature stop in the sequence SEQ ID NO:72 (Miller et al. J Biol Chem 2000; 275:32052-6). Additionally, the alignment of the protein of SEQ ID NO:248 with occludin, an integral membrane protein found at tight junctions (Furuse et al. J Cell Biol 1994; 127:1617-26), reveals that both proteins display a C-terminal ZO-1 binding domain, with a 26% homology over a 108 amino acid segment. Protein SEQ ID NO:313 lacks this domain, as its C-terminal region is truncated as compared to the protein of SEQ ID NO:248. ZO-1 is part of the family of membrane-associated guanylate kinase homologs (MAGUKs) believed to be important in signal transduction originating from sites of cell-cell contact (Willott et al. Proc Natl Acad Sci USA 1993; 90:7834-8).

The proteins of SEQ ID NOs:248 and 313 are RNA polymerase II elongation factors that increase the catalytic rate of transcription elongation, a phase during which RNA polymerase II moves along the DNA and extends the growing RNA chain (Miller et al. J Biol Chem 2000;

275:32052-6). Specifically, the proteins of SEQ ID NOs:248 and 313 suppress transient pausing at multiple sites along the DNA, thereby altering the K_m and/or the V_{max} of the polymerase (Miller et al. J Biol Chem 2000; 275:32052-6). The present proteins belong to a family that is known to include one virally encoded protein (Tat) and six cellular proteins (SIX, P-TEFb, TFIIF, Elongin 5 (SIII), ELL and ELL2).

A growing body of evidence suggests that mis-regulation of elongation may be a key element in a variety of human diseases (see, Aso et al. J Clin Invest 1996; 97:1561-9). For instance, two RNA polymerase II elongation proteins have been implicated in oncogenesis: ELL, which is a frequent target for translocation in acute myeloid leukemia (Thirman et al. Proc Natl Acad Sci USA 10 1994; 91:12110-4 ; Mitani et al. Blood 1995;85:2017-24), and elongin, which is a transcription factor regulated by the product of the von Hippel-Lindau tumor suppressor gene, which is itself mutated in the majority of clear-cell renal carcinomas and in families with von Hippel-Lindau disease (Duan et al. Science 1995;269:1402-6, Kibel et al. Science 1995; 269:1444-6). In addition, overexpression of ELL leads to the transformation of fibroblasts (Kanda et al. J Biol Chem. 1998 15 27; 273:5248-52). Thus, the proteins of SEQ ID NOs:248 and 313 may be important for oncogenesis of multiple types of neoplastic diseases, especially hematological malignancies.

In one embodiment, the present proteins are used to increase the rate of transcription in vitro. Such an increase can be used for any of the large number of in vitro transcription reactions which are routinely used for molecular biological applications, e.g. for the preparation of RNA, for 20 protein production, for the characterization of promoters and transcription factors, etc.

In another embodiment, the present invention provides diagnostic tools for the detection of mutations in the genes encoding SEQ ID NOs:248 or 313. Such mutations may be detected by a variety of techniques, including RNase and S1 protection assays; alterations in electrophoretic mobility of DNA fragments in gels, with or without denaturing agents such as SSCP or DGGE; 25 dHPLC; and direct DNA sequencing. The detection of mutations in the genes encoding SEQ ID NOs:248 or 313 are useful for the detection of a number of diseases and conditions, such as cancers and hematological malignancies including leukemia. For example, the RNA polymerase II Elongation Factor ELL gene undergoes frequent translocations in acute myeloid leukemia (Thirman et al. Proc Natl Acad Sci USA 1994; 91:12110-4 ; Mitani et al. Blood 1995; 85:2017-24), and it is 30 likely that other elongation factors are involved in additional such diseases.

Another embodiment of the present inventions relates to compositions and methods for using the proteins or part thereof to specifically visualize myeloid precursor cells, as well as pancreas, liver and testis tissues (and cells derived from the tissues). The ability to detect such cell types is useful for any of a number of applications, including to determine the origin or identity of, 35 e.g. cancerous cells, as well as to facilitate the identification of particular cells and tissues for, e.g. the evaluation of histological slides. In addition, such methods can be used to examine the extent of differentiation in myeloid or myeloid-progenitor cells for staging of leukemia or any other

neoplastic disorder. Any method for detecting the presence of the proteins of the invention, or nucleic acids encoding the proteins, can be used, including methods involving the use of antibodies immunospecific for the proteins of invention. Such antibodies can be used in various methods including radioimmunoassays, competitive binding assays, Western Blot analysis and enzyme-linked immunosorbant assay (ELISA) assays, or any other technique known to those skilled in the art. In another embodiment, the present protein or part thereof can be used for the treatment, attenuation and/or prevention of conditions associated with unbalanced amounts and/or activity of the protein of SEQ ID NO:248 or 313. Other modulatory substances can also be used in such embodiments, including chemical compounds such as agonists and antagonists, nucleic acids including antisense and ribozyme sequences, and antibodies. In a preferred embodiment, such substances are employed for the treatment or prevention of certain types of neoplastic disorders associated such as cancer or hematological malignancies such as leukemia. In such embodiments, where an increased level of expression or activity of the present proteins is correlated with the presence of a disease such as cancer, the disease can be treated or prevented using any agent that can provoke a decrease in the level of activity or expression of the protein, such as antibodies, antisense molecules, ribozymes, dominant negative forms of the protein, compounds that inhibit the expression or activity of the proteins, and others. Alternatively, in cases where a decreased level of expression or activity of the proteins is correlated with the presence of a disease such as cancer, the disease can be treated using any agent that can cause an increase in the expression or activity of the protein, such as polynucleotides encoding the proteins, purified forms of the proteins, or any compound that causes an increase in the expression or activity of the proteins. Further, any detection of a correlation between the level of expression or activity of the protein and the presence or absence of a disease can be used to develop diagnostic or screening tools for the detection of the disease itself, or of a predisposition for the disease.

25 Uses of antibodies

Antibodies of the present invention have uses that include, but are not limited to, methods known in the art to purify, detect, and target the polypeptides of the present invention including both *in vitro* and *in vivo* diagnostic and therapeutic methods. An example of such use using immunoaffinity chromatography is given below. The antibodies of the present invention may be used either alone or in combination with other compositions. For example, the antibodies have use in immunoassays for qualitatively and quantitatively measuring levels of antigen-bearing substances, including the polypeptides of the present invention, in biological samples (*See, e.g., Harlow et al., 1988*). (Incorporated by reference in the entirety). The antibodies may also be used in therapeutic compositions for killing cells expressing the protein or reducing the levels of the protein in the body.

The invention further relates to antibodies that act as agonists or antagonists of the polypeptides of the present invention. For example, the present invention includes antibodies that

disrupt the receptor/ligand interactions with the polypeptides of the invention either partially or fully. Included are both receptor-specific antibodies and ligand-specific antibodies. Included are receptor-specific antibodies, which do not prevent ligand binding but prevent receptor activation. Receptor activation (i.e., signaling) may be determined by techniques described herein or otherwise known in the art. Also included are receptor-specific antibodies which both prevent ligand binding and receptor activation. Likewise, included are neutralizing antibodies that bind the ligand and prevent binding of the ligand to the receptor, as well as antibodies that bind the ligand, thereby preventing receptor activation, but do not prevent the ligand from binding the receptor. Further included are antibodies that activate the receptor. These antibodies may act as agonists for either all or less than all of the biological activities affected by ligand-mediated receptor activation. The antibodies may be specified as agonists or antagonists for biological activities comprising specific activities disclosed herein. The above antibody agonists can be made using methods known in the art. *See e.g.*, WO 96/40281; US Patent 5,811,097; Deng *et al.* (1998); Chen *et al.* (1998); Harrop *et al.* (1998); Zhu *et al.* (1998); Yoon *et al.* (1998); Prat *et al.* (1998); Pitard *et al.* (1997); Liautard *et al.* (1997); Carlson *et al.* (1997); Taryman *et al.* (1995); Muller *et al.* (1998); Bartunek *et al.* (1996) (said references incorporated by reference in their entireties).

As discussed above, antibodies of the polypeptides of the invention can, in turn, be utilized to generate anti-idiotypic antibodies that "mimic" polypeptides of the invention using techniques well known to those skilled in the art (*See, e.g.* Greenspan and Bona (1989) and Nissinoff (1991), which disclosures are hereby incorporated by reference in their entireties). For example, antibodies which bind to and competitively inhibit polypeptide multimerization or binding of a polypeptide of the invention to ligand can be used to generate anti-idiotypes that "mimic" the polypeptide multimerization or binding domain and, as a consequence, bind to and neutralize polypeptide or its ligand. Such neutralization anti-idiotypic antibodies can be used to bind a polypeptide of the invention or to bind its ligands/receptors, and thereby block its biological activity.

Immunoaffinity Chromatography

Antibodies prepared as described herein are coupled to a support. Preferably, the antibodies are monoclonal antibodies, but polyclonal antibodies may also be used. The support may be any of those typically employed in immunoaffinity chromatography, including Sepharose CL-4B (Pharmacia, Piscataway, NJ), Sepharose CL-2B (Pharmacia, Piscataway, NJ), Affi-gel 10 (Biorad, Richmond, CA), or glass beads.

The antibodies may be coupled to the support using any of the coupling reagents typically used in immunoaffinity chromatography, including cyanogen bromide. After coupling the antibody to the support, the support is contacted with a sample which contains a target polypeptide whose isolation, purification or enrichment is desired. The target polypeptide may be a polypeptide selected from the group consisting of sequences of SEQ ID Nos: 242-482, mature polypeptides

included in SEQ ID Nos: 242-272 and 274-384 as well as full-length and mature polypeptides encoded by the clone inserts of the deposited clone pool, variants and fragments thereof, or a fusion protein comprising said selected polypeptide or a fragment thereof.

Preferably, the sample is placed in contact with the support for a sufficient amount of time
5 and under appropriate conditions to allow at least 50% of the target polypeptide to specifically bind to the antibody coupled to the support.

Thereafter, the support is washed with an appropriate wash solution to remove polypeptides which have non-specifically adhered to the support. The wash solution may be any of those typically employed in immunoaffinity chromatography, including PBS, Tris-lithium chloride buffer
10 (0.1M lysine base and 0.5M lithium chloride, pH 8.0), Tris-hydrochloride buffer (0.05M Tris-hydrochloride, pH 8.0), or Tris/Triton/NaCl buffer (50mM Tris.cl, pH 8.0 or 9.0, 0.1% Triton X-100, and 0.5MNaCl).

After washing, the specifically bound target polypeptide is eluted from the support using the high pH or low pH elution solutions typically employed in immunoaffinity chromatography. In
15 particular, the elution solutions may contain an eluant such as triethanolamine, diethylamine, calcium chloride, sodium thiocyanate, potassium bromide, acetic acid, or glycine. In some embodiments, the elution solution may also contain a detergent such as Triton X-100 or octyl-beta-D-glucoside.

Import vectors

20 The GENSET polypeptides of the invention may also be used as a carrier to import a protein or peptide of interest, so-called cargo, into tissue-culture cells or in host organisms. A hydrophobic region of a GENSET polypeptide or a fragment thereof, preferably the signal peptide of a sequence selected from the group consisting of of SEQ ID Nos: 1-31 and 33-143 and clones inserts of the deposited clone pool, more preferably the short core hydrophobic region (h) of signal
25 peptides may be used as a carrier.

When cell permeable peptides of limited size (approximately up to 25 amino acids) are to be translocated across cell membrane, chemical synthesis may be used in order to add the h region to either the C-terminus or the N-terminus to the cargo peptide of interest. Alternatively, when longer peptides or proteins are to be imported into cells, nucleic acids can be genetically engineered,
30 using techniques familiar to those skilled in the art, in order to link the cDNA sequence or fragment thereof encoding the hydrophobic region to the 5' or the 3' end of a DNA sequence coding for a cargo polypeptide. Such genetically engineered nucleic acids are then translated either *in vitro* or *in vivo* after transfection into appropriate cells, using conventional techniques to produce the resulting cell permeable polypeptide. Suitable hosts cells are then simply incubated with the cell permeable
35 polypeptide which is then translocated across the membrane.

This method may be applied to study diverse intracellular functions and cellular processes. For instance, it has been used to probe functionally relevant domains of intracellular proteins and to examine protein-protein interactions involved in signal transduction pathways (Lin *et al.*, *J. Biol. Chem.*, 270: 14225-14258 (1995); Lin *et al.*, *J. Biol. Chem.*, 271: 5305-5308 (1996); Rojas *et al.*, *J. Biol. Chem.*, 271: 27456-27461 (1996); Rojas *et al.*, *Nature Biotech.*, 16: 370-375 (1998); Liu *et al.*, *Proc. Natl. Acad. Sci. USA*, 93: 11819-11824 (1996); Rojas *et al.*, *Bioch. Biophys. Res. Commun.*, 234: 675-680 (1997) Du *et al.*, *J. Peptide Res.*, 51: 235-243 (1998)).

Such techniques may be used in cellular therapy to import proteins producing therapeutic effects. For instance, cells isolated from a patient may be treated with imported therapeutic proteins and then re-introduced into the host organism.

Alternatively, the hydrophobic region of signal peptides of the present invention could be used in combination with a nuclear localization signal to deliver nucleic acids into cell nucleus. Such oligonucleotides may be antisense oligonucleotides or oligonucleotides designed to form triple helixes, as described herein, in order to respectively inhibit processing or maturation of a target cellular RNA.

EXPRESSION OF GENSET PRODUCTS

Spatial expression of the GENSET genes of the invention

Tissue expression of the cDNAs of the present invention was examined. Table IX list the Genset's libraries of tissues and cell types examined that express the polynucleotides of the present invention. The tissues and cell types examined for polynucleotide expression were: adrenal gland (AG), bone marrow (BM), brain (Br), cancerous prostate (CP), cerebellum (Ce), colon (Co), dystrophic muscle (DM), fetal brain (FB), fetal kidney (FK), fetal liver (FL), heart (He), hypertrophic prostate (HP), kidney (Ki), liver (Li), lung (Lu), lung cells (LC), lymph ganglia (LG), lymphocytes (Ly), muscle (Mu), Ovary (Ov), pancreas (Pa), pituitary gland (PG), placenta (Pl), prostate (Pr), salivary gland (SG), spinal cord (SC), spleen (Sp), stomach/intestine (SI), substantia nigra (SN), testis (Te), thyroid (Ty), umbilical cord (UC) and uterus (Ut).

For each cDNA referred to by its sequence identification number (first column), the number of proprietary 5'ESTs (i.e. cDNA fragments) expressed in a particular tissue referred to by its name is indicated after a semi column (second column). In addition, the bias in the spatial distribution of the polynucleotide sequences of the present invention was examined by comparing the relative proportions of the biological polynucleotides of a given tissue using the following statistical analysis. The under- or over-representation of a polynucleotide of a given cluster in a given tissue was performed using the normal approximation of the binomial distribution. When the observed proportion of a polynucleotide of a given tissue in a given consensus had less than 1% chance to occur randomly according to the chi2 test, the frequency bias was reported as "preferred". The results are given in Table X as follows. For each polynucleotide showing a bias in tissue distribution

as referred to by its sequence identification number in the first column, the list of tissues where the polynucleotides are under-represented is given in the second column entitled "low frequency expression" and the list of tissues where the polynucleotides are over-represented is given in the third column entitled "high frequency expression".

5 The cellular localization of some polypeptides of the invention was also determined using the "psort software" (Nakai, and Horton, (1999); Nakai and Kanehisa, (1992), which disclosures are hereby incorporated by reference in their entireties). For each polypeptide identified by its sequence identification number in the first column, the second column of Table XI list the predicted subcellular localization.

10 Evaluation of Expression Levels and Patterns of GENSET mRNAs

The spatial and temporal expression patterns of GENSET mRNAs, as well as their expression levels, may also be further determined as follows.

Expression levels and patterns of GENSET mRNAs may be analyzed by solution hybridization with long probes as described in International Patent Application No. WO 97/05277,
15 the entire contents of which are hereby incorporated by reference. Briefly, a GENSET polynucleotide, or fragment thereof corresponding to the gene encoding the mRNA to be characterized is inserted at a cloning site immediately downstream of a bacteriophage (T3, T7 or SP6) RNA polymerase promoter to produce antisense RNA. Preferably, the GENSET polynucleotide is at least a 100 nucleotides in length. The plasmid is linearized and transcribed in
20 the presence of ribonucleotides comprising modified ribonucleotides (i.e. biotin-UTP and DIG-UTP). An excess of this doubly labeled RNA is hybridized in solution with mRNA isolated from cells or tissues of interest. The hybridizations are performed under standard stringent conditions (40-50°C for 16 hours in an 80% formamide, 0.4 M NaCl buffer, pH 7-8). The unhybridized probe is removed by digestion with ribonucleases specific for single-stranded RNA (i.e. RNases CL3, T1,
25 Phy M, U2 or A). The presence of the biotin-UTP modification enables capture of the hybrid on a microtitration plate coated with streptavidin. The presence of the DIG modification enables the hybrid to be detected and quantified by ELISA using an anti-DIG antibody coupled to alkaline phosphatase.

The GENSET cDNAs, or fragments thereof may also be tagged with nucleotide sequences
30 for the serial analysis of gene expression (SAGE) as disclosed in UK Patent Application No. 2 305 241 A, the entire contents of which are incorporated by reference. In this method, cDNAs are prepared from a cell, tissue, organism or other source of nucleic acid for which it is desired to determine gene expression patterns. The resulting cDNAs are separated into two pools. The cDNAs in each pool are cleaved with a first restriction endonuclease, called an "anchoring enzyme,"
35 having a recognition site which is likely to be present at least once in most cDNAs. The fragments which contain the 5' or 3' most region of the cleaved cDNA are isolated by binding to a capture

medium such as streptavidin coated beads. A first oligonucleotide linker having a first sequence for hybridization of an amplification primer and an internal restriction site for a "tagging endonuclease" is ligated to the digested cDNAs in the first pool. Digestion with the second endonuclease produces short "tag" fragments from the cDNAs. A second oligonucleotide having a second sequence for hybridization of an amplification primer and an internal restriction site is ligated to the digested cDNAs in the second pool. The cDNA fragments in the second pool are also digested with the "tagging endonuclease" to generate short "tag" fragments derived from the cDNAs in the second pool. The "tags" resulting from digestion of the first and second pools with the anchoring enzyme and the tagging endonuclease are ligated to one another to produce "ditags." In some embodiments, the ditags are concatamerized to produce ligation products containing from 2 to 200 ditags. The tag sequences are then determined and compared to the sequences of the GENSET cDNAs to determine which genes are expressed in the cell, tissue, organism, or other source of nucleic acids from which the tags were derived. In this way, the expression pattern of a GENSET gene in the cell, tissue, organism, or other source of nucleic acids is obtained.

Quantitative analysis of GENSET gene expression may also be performed using arrays. For example, quantitative analysis of gene expression may be performed with GENSET polynucleotides, or fragments thereof in a complementary DNA microarray as described by Schena *et al.* (1995 and 1996) which disclosures are hereby incorporated by reference in their entireties. GENSET cDNAs or fragments thereof are amplified by PCR and arrayed from 96-well microtiter plates onto silylated microscope slides using high-speed robotics. Printed arrays are incubated in a humid chamber to allow rehydration of the array elements and rinsed, once in 0.2% SDS for 1 min, twice in water for 1 min and once for 5 min in sodium borohydride solution. The arrays are submerged in water for 2 min at 95°C, transferred into 0.2% SDS for 1 min, rinsed twice with water, air dried and stored in the dark at 25°C. Cell or tissue mRNA is isolated or commercially obtained and probes are prepared by a single round of reverse transcription. Probes are hybridized to 1 cm² microarrays under a 14 x 14 mm glass coverslip for 6-12 hours at 60°C. Arrays are washed for 5 min at 25°C in low stringency wash buffer (1X SSC/0.2% SDS), then for 10 min at room temperature in high stringency wash buffer (0.1X SSC/0.2% SDS). Arrays are scanned in 0.1X SSC using a fluorescence laser scanning device fitted with a custom filter set. Accurate differential expression measurements are obtained by taking the average of the ratios of two independent hybridizations.

Quantitative analysis of the expression of genes may also be performed with GENSET cDNAs or fragments thereof in complementary DNA arrays as described by Pietu *et al.* (1996), which disclosure is hereby incorporated by reference in its entirety. The GENSET polynucleotides of the invention or fragments thereof are PCR amplified and spotted on membranes. Then, mRNAs originating from various tissues or cells are labeled with radioactive nucleotides. After hybridization and washing in controlled conditions, the hybridized mRNAs are detected by

phospho-imaging or autoradiography. Duplicate experiments are performed and a quantitative analysis of differentially expressed mRNAs is then performed.

Alternatively, expression analysis of GENSET genes can be done through high density nucleotide arrays as described by Lockhart *et al.* (1996) and Sosnowski *et al.* (1997), which disclosures are hereby incorporated by reference in their entireties. Oligonucleotides of 15-50 nucleotides corresponding to sequences of a GENSET polynucleotide or fragments thereof are synthesized directly on the chip (Lockhart *et al.*, supra) or synthesized and then addressed to the chip (Sosnowski *et al.*, supra). Preferably, the oligonucleotides are about 20 nucleotides in length. cDNA probes labeled with an appropriate compound, such as biotin, digoxigenin or fluorescent dye, are synthesized from the appropriate mRNA population and then randomly fragmented to an average size of 50 to 100 nucleotides. The said probes are then hybridized to the chip. After washing as described in Lockhart *et al.*, (supra) and application of different electric fields (Sosnowski *et al.*, supra), the dyes or labeling compounds are detected and quantified. Duplicate hybridizations are performed. Comparative analysis of the intensity of the signal originating from cDNA probes on the same target oligonucleotide in different cDNA samples indicates a differential expression of the GENSET mRNA.

Uses of GENSET expression data

Once the expression levels and patterns of a GENSET mRNA has been determined using any technique known to those skilled in the art, in particular those described in the section entitled “Evaluation of Expression Levels and Patterns of GENSET mRNAs”, or using the instant disclosure, these information may be used to design GENSET specific markers for detection, identification, screening and diagnosis purposes as well as to design DNA constructs with an expression pattern similar to a GENSET expression pattern.

Detection of GENSET expression and/or biological activity

The invention further relates to methods of detection of GENSET expression and/or biological activity in a biological sample using the polynucleotide and polypeptide sequences described herein. Such method scan be used, for example, as a screen for normal or abnormal GENSET expression and/or biological activity and, thus, can be used diagnostically. The biological sample for use in the methods of the present invention includes a suitable sample from, for example, a mammal, particularly a human. For example, the sample can be issued from tissues or cell lines having the same origin as tissues or cell lines in which the polypeptide is known to be expressed using the data from Table IX.

Detection of GENSET products

The invention further relates to methods of detection of GENSET polynucleotides or polypeptides in a sample using the sequences described herein and any techniques known to those

skilled in the art. For example, a labeled polynucleotide probe having all or a functional portion of the nucleotide sequence of a GENSET polynucleotide can be used in a method to detect a GENSET polynucleotide in a sample. In one embodiment, the sample is treated to render the polynucleotides in the sample available for hybridization to a polynucleotide probe, which can be DNA or RNA.

- 5 The resulting treated sample is combined with a labeled polynucleotide probe having all or a portion of the nucleotide sequence of the GENSET cDNA or genomic sequence, under conditions appropriate for hybridization of complementary sequences to occur. Detection of hybridization of polynucleotides from the sample with the labeled nucleic probe indicates the presence of GENSET polynucleotides in a sample. The presence of GENSET mRNA is indicative of GENSET
- 10 expression.

Consequently, the invention comprises methods for detecting the presence of a polynucleotide comprising a nucleotide sequence selected from a group consisting of the sequences of SEQ ID Nos: 1-241, the sequences of clone inserts of the deposited clone pool, sequences fully complementary thereto, fragments and variants thereof in a sample. In a first embodiment, said

15 method comprises the following steps of:

a) bringing into contact said sample and a nucleic acid probe or a plurality of nucleic acid probes which hybridize to said selected nucleotide sequence; and

b) detecting the hybrid complex formed between said probe or said plurality of probes and said polynucleotide.

- 20 In a preferred embodiment of the above detection method, said nucleic acid probe or said plurality of nucleic acid probes is labeled with a detectable molecule. In another preferred embodiment of the above detection method, said nucleic acid probe or said plurality of nucleic acid probes has been immobilized on a substrate. In still another preferred embodiment, said nucleic acid probe or said plurality of nucleic acid probes has a sequence comprised in a sequence
- 25 complementary to said selected sequence.

In a second embodiment, said method comprises the following steps of:

a) contacting said sample with amplification reaction reagents comprising a pair of amplification primers located on either side of the region of said nucleotide sequence to be amplified;

- 30 b) performing an amplification reaction to synthesize amplification products containing said region of said selected nucleotide sequence; and

c) detecting said amplification products.

- In a preferred embodiment of the above detection method, when the polynucleotide to be amplified is a RNA molecule, preliminary reverse transcription and synthesis of a second cDNA
- 35 strand are necessary to provide a DNA template to be amplified. In another preferred embodiment of the above detection method, the amplification product is detected by hybridization with a labeled probe having a sequence which is complementary to the amplified region. In still another preferred

embodiment, at least one of said amplification primer has a sequence comprised in said selected sequence or in the sequence complementary to said selected sequence.

Alternatively, a method of detecting GENSET expression in a test sample can be accomplished using any product which binds to a GENSET polypeptide of the present invention or
5 a portion of a GENSET polypeptide. Such products may be antibodies, binding fragments of antibodies, polypeptides able to bind specifically to GENSET polypeptides or fragments thereof, including GENSET agonists and antagonists. Detection of specific binding to the antibody indicates the presence of a GENSET polypeptide in the sample (e.g., ELISA).

Consequently, the invention is also directed to a method for detecting specifically the
10 presence of a GENSET polypeptide according to the invention in a biological sample, said method comprising the following steps of:

a) bringing into contact said biological sample with a product able to bind to a polypeptide of the invention or fragments thereof;

b) allowing said product to bind to said polypeptide to form a complex; and

15 b) detecting said complex.

In a preferred embodiment of the above detection method, the product is an antibody. In a more preferred embodiment, said antibody is labeled with a detectable molecule. In another more preferred embodiment of the above detection method, said antibody has been immobilized on a substrate.

20 In addition, the invention also relates to methods of determining whether a GENSET product (e.g. a polynucleotide or polypeptide) is present or absent in a biological sample, said methods comprising the steps of:

a) obtaining said biological sample from a human or non-human animal, preferably a mammal;

25 b) contacting said biological sample with a product able to bind to a GENSET polynucleotide or polypeptide of the invention; and

c) determining the presence or absence of said GENSET product in said biological sample.

Compounds that specifically binds a GENSET product may either be compounds binding to a GENSET polypeptide (e.g. binding proteins, antibodies or binding fragments thereof (e.g. F(ab')₂
30 fragments) or compounds bindint to a GENSET polynucleotide (e.g. a complementary probe or primer).

The present invention also relates to kits that can be used in the detection of GENSET expression products. The kit can comprise a compound that specifically binds a GENSET polypeptide (e.g. binding proteins, antibodies or binding fragments thereof (e.g. F(ab')₂ fragments)
35 or a GENSET mRNA (e.g. a complementary probe or primer), for example, disposed within a container means. The kit can further comprise ancillary reagents, including buffers and the like.

Detection of a GENSET biological activity

The invention further includes methods of detecting specifically a GENSET biological activity. Assessing the GENSET biological activity may be performed using a variety of techniques, including those described in the section entitled “**Erreur! Source du renvoi introuvable.**”.

- 5 Consequently, the invention is directed to a method for detecting specifically GENSET biological activity in a biological sample, said method comprising the following steps:
- a) obtaining a biological sample from a human or non-human mammal; and
 - b) detecting a GENSET biological activity.

 The present invention also relates to kits that can be used in the detection of GENSET
10 biological activity.

Identification of a specific context of GENSET expression

 When the expression pattern of a GENSET mRNA shows that a GENSET gene is specifically expressed in a given context, probes and primers specific for this gene as well as antibodies binding to the GENSET polynucleotide may then be used as markers for a specific
15 context. Examples of specific contexts are: specific expression in a given tissue/cell or tissue/cell type, expression at a given stage of development of a process such as embryo development or disease development, or specific expression in a given organelle. Such primers, probes, and antibodies are useful commercially to identify tissues/cells/organelles of unknown origin, for example, forensic samples, differentiated tumor tissue that has metastasized to foreign bodily sites,
20 or to differentiate different tissue types in a tissue cross-section using any technique known to those skilled in the art including *in situ* PCR or immunochemistry for example.

 For example, the cDNAs and proteins of the sequence listing and fragments thereof, may be used to distinguish human tissues/cells from non-human tissues/cells and to distinguish between human tissues/cells/organelles that do and do not express the polynucleotides comprising the
25 cDNAs. By knowing the expression pattern of a given GENSET, either through routine experimentation or by using the instant disclosure, the polynucleotides and polypeptides of the present invention may be used in methods of determining the identity of an unknown tissue/cell sample/organelle. As part of determining the identity of an unknown tissue/cell sample/organelle, the polynucleotides and polypeptides of the present invention may be used to determine what the
30 unknown tissue/cell sample is and what the unknown sample is not. For example, if a cDNA is expressed in a particular tissue/cell type/organelle, and the unknown tissue/cell sample/organelle does not express the cDNA, it may be inferred that the unknown tissue/cells are either not human or not the same human tissue/cell type/organelle as that which expresses the cDNA. These methods of determining tissue/cell/organelle identity are based on methods which detect the presence or
35 absence of the mRNA (or corresponding cDNA) in a tissue/cell sample using methods well known in the art (e.g., hybridization, PCR based methods, immunoassays, immunochemistry, ELISA).

Examples of such techniques are described in more detail below. Therefore, the invention encompasses uses of the polynucleotides and polypeptides of the invention as tissue markers. In a preferred embodiment, polynucleotides preferentially expressed in given tissues as indicated in Table X and polypeptides encoded by such polynucleotides are used for this purpose. The invention also encompasses uses of polypeptides of the invention as organelle markers. In a preferred embodiment, polypeptides preferentially expressed in given subcellular compartment as indicated in Table XI are used for this purpose.

Consequently, the present invention encompasses methods of identification of a tissue/cell type/subcellular compartment, wherein said method includes the steps of:

- a) contacting a biological sample which identity is to be assayed with a product able to bind a GENSET product; and
- b) determining whether a GENSET product is expressed in said biological sample.

Products that are able to bind specifically to a GENSET product, namely a GENSET polypeptide or a GENSET mRNA, include GENSET binding proteins, antibodies or binding fragments thereof (e.g. F(ab')₂ fragments), as well as GENSET complementary probes and primers.

Step b) may be performed using any detection method known to those skilled in the art including those disclosed herein, especially in the section entitled "Detection of GENSET expression and/or biological activity"..

20 *Identification of Tissue Types or Cell Species by Means of Labeled Tissue Specific Antibodies*

Identification of specific tissues is accomplished by the visualization of tissue specific antigens by means of antibody preparations which are conjugated, directly (e.g., green fluorescent protein) or indirectly to a detectable marker. Selected labeled antibody species bind to their specific antigen binding partner in tissue sections, cell suspensions, or in extracts of soluble proteins from a tissue sample to provide a pattern for qualitative or semi-qualitative interpretation.

Antisera for these procedures must have a potency exceeding that of the native preparation, and for that reason, antibodies are concentrated to a mg/ml level by isolation of the gamma globulin fraction, for example, by ion-exchange chromatography or by ammonium sulfate fractionation. Also, to provide the most specific antisera, unwanted antibodies, for example to common proteins, must be removed from the gamma globulin fraction, for example by means of insoluble immunoabsorbents, before the antibodies are labeled with the marker. Either monoclonal or heterologous antisera is suitable for either procedure.

A. Immunohistochemical Techniques

Purified, high-titer antibodies, prepared as described above, are conjugated to a detectable marker, as described, for example, by Fudenberg, (1980) or Rose *et al.*, (1980), which disclosures are hereby incorporated by reference in their entireties.

A fluorescent marker, either fluorescein or rhodamine, is preferred, but antibodies can also be labeled with an enzyme that supports a color producing reaction with a substrate, such as horseradish peroxidase. Markers can be added to tissue-bound antibody in a second step, as described below. Alternatively, the specific anti-tissue antibodies can be labeled with ferritin or other electron dense particles, and localization of the ferritin coupled antigen-antibody complexes achieved by means of an electron microscope. In yet another approach, the antibodies are radiolabeled, with, for example ^{125}I , and detected by overlaying the antibody treated preparation with photographic emulsion. Preparations to carry out the procedures can comprise monoclonal or polyclonal antibodies to a single protein or peptide identified as specific to a tissue type, for example, brain tissue, or antibody preparations to several antigenically distinct tissue specific antigens can be used in panels, independently or in mixtures, as required. Tissue sections and cell suspensions are prepared for immunohistochemical examination according to common histological techniques. Multiple cryostat sections (about 4 μm , unfixed) of the unknown tissue and known control, are mounted and each slide covered with different dilutions of the antibody preparation. Sections of known and unknown tissues should also be treated with preparations to provide a positive control, a negative control, for example, pre-immune sera, and a control for non-specific staining, for example, buffer. Treated sections are incubated in a humid chamber for 30 min at room temperature, rinsed, then washed in buffer for 30-45 min. Excess fluid is blotted away, and the marker developed. If the tissue specific antibody was not labeled in the first incubation, it can be labeled at this time in a second antibody-antibody reaction, for example, by adding fluorescein- or enzyme-conjugated antibody against the immunoglobulin class of the antiserum-producing species, for example, fluorescein labeled antibody to mouse IgG. Such labeled sera are commercially available. The antigen found in the tissues by the above procedure can be quantified by measuring the intensity of color or fluorescence on the tissue section, and calibrating that signal using appropriate standards.

B. Identification of Tissue Specific Soluble Proteins

The visualization of tissue specific proteins and identification of unknown tissues from that procedure is carried out using the labeled antibody reagents and detection strategy as described for immunohistochemistry; however the sample is prepared according to an electrophoretic technique to distribute the proteins extracted from the tissue in an orderly array on the basis of molecular weight for detection. A tissue sample is homogenized using a Virtis apparatus; cell suspensions are disrupted by Dounce homogenization or osmotic lysis, using detergents in either case as required to disrupt cell membranes, as is the practice in the art. Insoluble cell components such as nuclei, microsomes, and membrane fragments are removed by ultracentrifugation, and the soluble protein-containing fraction concentrated if necessary and reserved for analysis. A sample of the soluble protein solution is resolved into individual protein species by conventional SDS polyacrylamide

electrophoresis as described, for example, by Davis *et al.*, Section 19-2 (1986), using a range of amounts of polyacrylamide in a set of gels to resolve the entire molecular weight range of proteins to be detected in the sample. A size marker is run in parallel for purposes of estimating molecular weights of the constituent proteins. Sample size for analysis is a convenient volume of from 5 to 55
5 ul, and containing from about 1 to 100 ug protein. An aliquot of each of the resolved proteins is transferred by blotting to a nitrocellulose filter paper, a process that maintains the pattern of resolution. Multiple copies are prepared. The procedure, known as Western Blot Analysis, is well described in Davis *et al.*, (1986) Section 19-3. One set of nitrocellulose blots is stained with Coomassie Blue dye to visualize the entire set of proteins for comparison with the antibody bound
10 proteins. The remaining nitrocellulose filters are then incubated with a solution of one or more specific antisera to tissue specific proteins prepared as described herein. In this procedure, as in procedure A above, appropriate positive and negative sample and reagent controls are run.

In either procedure A or B, a detectable label can be attached to the primary tissue antigen-
primary antibody complex according to various strategies and permutations thereof. In a
15 straightforward approach, the primary specific antibody can be labeled; alternatively, the unlabeled complex can be bound by a labeled secondary anti-IgG antibody. In other approaches, either the primary or secondary antibody is conjugated to a biotin molecule, which can, in a subsequent step, bind an avidin conjugated marker. According to yet another strategy, enzyme labeled or radioactive protein A, which has the property of binding to any IgG, is bound in a final step to either the
20 primary or secondary antibody. The visualization of tissue specific antigen binding at levels above those seen in control tissues to one or more tissue specific antibodies, prepared from the gene sequences identified from cDNA sequences, can identify tissues of unknown origin, for example, forensic samples, or differentiated tumor tissue that has metastasized to foreign bodily sites.

Targeting of compounds to subcellular compartments

25 GENSET Polypeptides expressed in specific cellular compartments/organelles may also be used to target compounds to these compartments/organelles. The invention therefore encompasses uses of polypeptides and polynucleotides of the invention as organelle targeting tools.

In a first embodiment, GENSET polypeptides expressed in mitochondria may be used to target heterologous compounds, either polypeptides or polynucleotides to mitochondria by
30 recombinantly or chemically fusing a fragment of the protein of the invention to an heterologous polypeptide or polynucleotide. Preferred fragments are signal peptide, amphiphilic alpha helices and/or any other fragments of the protein of the invention, or part thereof, that may contain targeting signals for mitochondria including but not limited to matrix targeting signals as defined in Herrman and Neupert, (2000); Bhagwat *et al.* (1999), Murphy (1997); Glaser *et al.* (1998);
35 Ciminale *et al.* (1999), which disclosures are hereby incorporated by reference in their entireties. Such heterologous compounds may be used to modulate mitochondria's activities. For example,

they may be used to induce and/or prevent mitochondrial-induced apoptosis or necrosis. In addition, heterologous polynucleotides may be used for mitochondrial gene therapy to replace a defective mitochondrial gene and/or to inhibit the deleterious expression of a mitochondrial gene.

In a second embodiment, GENSET polypeptides expressed in the endoplasmic reticulum may be used to target heterologous polypeptides to the endoplasmic reticulum by recombinantly or chemically fusing a fragment of the proteins of the invention to an heterologous polypeptide. Preferred fragments are any fragments of the proteins of the invention, or part thereof, that may contain targeting signals for the endoplasmic reticulum such as those described in Pidoux and Armstrong (1992), Munro and Pelham (1987); Pelham (1990), which disclosures are hereby incorporated by reference in their entireties.

In a third embodiment, GENSET polypeptides expressed in the nucleus may be used to target heterologous polypeptides or polynucleotides to the nucleus by recombinantly or chemically fusing a fragment of the proteins or polynucleotide of the invention to an heterologous polypeptide or polynucleotide. Preferred fragments are any fragments of the proteins or polynucleotide of the invention, or part thereof, that may contain targeting signals for the nucleus (nuclear localization signals) such as those described in Christophe et al. (2000), which disclosure is hereby incorporated by reference in its entirety.

In a fourth embodiment, GENSET polypeptides expressed in the nucleus may be used to target heterologous polypeptides to the Golgi apparatus by recombinantly or chemically fusing a fragment of the protein of the invention to an heterologous polypeptide. Preferred fragments are signal peptide, transmembrane domains, tyrosine containing regions and/or any other fragments of the proteins of the invention, or part thereof, that may contain (1) targeting signals for the Golgi apparatus such as the ones described in Ugur and Jones, (2000); Picetti and Borrelli, (2000), (2) tyrosine-based Golgi targeting signal region (Zhan et al., (1998); Watson and Pessin (2000); Ward and Moss (2000), or (3) any other region as defined in Munro, (1998); Luetterforst et al., (1999); Essl et al., (1999), which disclosures are hereby incorporated by reference in their entireties.

Screening and diagnosis of abnormal GENSET expression and/or biological activity

Moreover, antibodies and/or primers specific for GENSET expression may also be used to identify abnormal GENSET expression and/or biological activity, and subsequently to screen and/or diagnose disorders associated with abnormal GENSET expression. For example, a particular disease may result from lack of expression, over expression, or under expression of a GENSET mRNA. By comparing mRNA expression patterns and quantities in samples taken from healthy individuals with those from individuals suffering from a particular disorder, genes responsible for this disorder may be identified. Primers, probes and antibodies specific for this GENSET may then be used to elaborate kits of screening and diagnosis for a disorder in which the gene of interest is

specifically expressed or in which its expression is specifically dysregulated, i.e. underexpressed or overexpressed.

Screening for specific disorders

The present invention also relates to methods of identifying individuals having elevated or
5 reduced levels of GENSET, which individuals are likely to benefit from therapies to suppress or enhance GENSET expression, respectively. One example of such methods comprises the steps of:

- a) obtaining from a human or non-human mammal a biological sample;
- b) detecting the presence in said sample of a GENSET product (mRNA or protein) using
any method known to those skilled in the art including those described herein, especially at the
10 section entitled "Detection of GENSET products";
- c) comparing the amount of said GENSET product present in said sample with that of a control sample; and
- d) determining whether said human or non-human mammal has a reduced or elevated level of GENSET expression compared to the control sample.

15 Such individuals with reduced or elevated levels of GENSET products may be predisposed to disorders associated with dysregulation of GENSET gene expression and thus would be candidates for therapies. The identification of elevated levels of GENSET in a patient would be indicative of an individual that would benefit from treatment with agents that suppress GENSET expression or activity. The identification of low levels of GENSET in a patient would be indicative
20 of an individual that would benefit from agents that induce GENSET expression or activity.

Biological samples suitable for use in this method include biological fluids such as blood, lymph, saliva, sperm, maternal milk, and tissue samples (e.g. biopsies) as well as cell cultures or cell extracts derived, for example, from tissue biopsies. The detection step of the present method can be performed using standard protocols for protein/mRNA detection. Examples of suitable
25 protocols include Northern blot analysis, immunoassays (e.g. RIA, Western blots, immunohistochemical analyses), and PCR.

Thus, the present invention further relates to methods of identifying individuals or non-human animals at increased risk for developing, or present state of having, certain diseases/disorders associated with GENSET abnormal expression or biological activity. One
30 example of such methods comprises the steps of:

- a) obtaining from a human or non-human mammal a biological sample;
- b) detecting the presence or absence in said sample of a GENSET product (mRNA or protein);
- c) comparing the amount of said GENSET product present in said sample with that of a
35 control sample; and

d) determining whether said human or non-human mammal is at increased risk for developing, or present state of having, a diseases or disorder.

In accordance with this method, the presence in the sample of altered levels of GENSET product indicates that the subject is predisposed to the above-indicated diseases/disorders.

- 5 Biological samples suitable for use in this method include biological fluids such as blood, lymph, saliva, sperm, maternal milk, and tissue samples (e.g. biopsies).

The diagnostic methodologies described herein are applicable to both humans and non-human mammals.

Detection of GENSET mutations

- 10 The invention also encompasses methods to detect mutations in GENSET polynucleotides of the invention. Such methods may advantageously be used to detect mutations occurring in GENSET genes and preferably in their regulatory regions. When the mutation was proven to be associated with a disease, screening for such mutations may be used for screening and diagnosis purposes.

- 15 In one embodiment of the oligonucleotide arrays of the invention, an oligonucleotide probe matrix may advantageously be used to detect mutations occurring in GENSET genes and preferably in their regulatory regions. For this particular purpose, probes are specifically designed to have a nucleotide sequence allowing their hybridization to the genes that carry known mutations (either by deletion, insertion or substitution of one or several nucleotides). By known mutations, it is meant,
20 mutations on the GENSET genes that have been identified according, for example to the technique used by Huang *et al.*(1996) or Samson *et al.*(1996), which disclosures are hereby incorporated by reference in their entireties.

- Another technique that is used to detect mutations in GENSET genes is the use of a high-density DNA array. Each oligonucleotide probe constituting a unit element of the high density
25 DNA array is designed to match a specific subsequence of a GENSET genomic DNA or cDNA. Thus, an array consisting of oligonucleotides complementary to subsequences of the target gene sequence is used to determine the identity of the target sequence with the wild gene sequence, measure its amount, and detect differences between the target sequence and the reference wild gene sequence of the GENSET gene. In one such design, termed 4L tiled array, is implemented a set of
30 four probes (A, C, G, T), preferably 15-nucleotide oligomers. In each set of four probes, the perfect complement will hybridize more strongly than mismatched probes. Consequently, a nucleic acid target of length L is scanned for mutations with a tiled array containing 4L probes, the whole probe set containing all the possible mutations in the known wild reference sequence. The hybridization signals of the 15-mer probe set tiled array are perturbed by a single base change in the target
35 sequence. As a consequence, there is a characteristic loss of signal or a "footprint" for the probes

flanking a mutation position. This technique was described by Chee *et al.* in 1996, which disclosure is hereby incorporated by reference in its entirety.

Construction of DNA constructs with a GENSET expression pattern

In addition, characterization of the spatial and temporal expression patterns and expression
5 levels of GENSET mRNAs is also useful for constructing expression vectors capable of producing a desired level of gene product in a desired spatial or temporal manner, as discussed below.

DNA Construct That Enables Directing Temporal And Spatial GENSET Gene Expression In Recombinant Cell Hosts And In Transgenic Animals.

In order to study the physiological and phenotypic consequences of a lack of synthesis of a
10 GENSET protein, both at the cell level and at the multi cellular organism level, the invention also encompasses DNA constructs and recombinant vectors enabling a conditional expression of a specific allele of a GENSET genomic sequence or cDNA and also of a copy of this genomic sequence or cDNA harboring substitutions, deletions, or additions of one or more bases as regards to a nucleotide sequence selected from the group consisting of sequences of SEQ ID Nos 1-241 and
15 sequences of clone inserts of the deposited clone pool, or a fragment thereof, these base substitutions, deletions or additions being located either in an exon, an intron or a regulatory sequence, but preferably in the 5'-regulatory sequence or in an exon of the GENSET genomic sequence or within the GENSET cDNA.

A first preferred DNA construct is based on the tetracycline resistance operon *tet* from *E.*
20 *coli* transposon Tn10 for controlling the GENSET gene expression, such as described by Gossen *et al.* (1992, 1995) and Furth *et al.* (1994), which disclosures are hereby incorporated by reference in their entireties. Such a DNA construct contains seven *tet* operator sequences from Tn10 (*tetop*) that are fused to either a minimal promoter or a 5'-regulatory sequence of the GENSET gene, said minimal promoter or said GENSET regulatory sequence being operably linked to a polynucleotide
25 of interest that codes either for a sense or an antisense oligonucleotide or for a polypeptide, including a GENSET polypeptide or a peptide fragment thereof. This DNA construct is functional as a conditional expression system for the nucleotide sequence of interest when the same cell also comprises a nucleotide sequence coding for either the wild type (tTA) or the mutant (rTA) repressor fused to the activating domain of viral protein VP16 of herpes simplex virus, placed under the
30 control of a promoter, such as the HCMVIE1 enhancer/promoter or the MMTV-LTR. Indeed, a preferred DNA construct of the invention comprise both the polynucleotide containing the *tet* operator sequences and the polynucleotide containing a sequence coding for the tTA or the rTA repressor. In a specific embodiment, the conditional expression DNA construct contains the sequence encoding the mutant tetracycline repressor rTA, the expression of the polynucleotide of
35 interest is silent in the absence of tetracycline and induced in its presence.

DNA Constructs Allowing Homologous Recombination: Replacement Vectors

A second preferred DNA construct will comprise, from 5'-end to 3'-end: (a) a first nucleotide sequence that is comprised in the GENSET genomic sequence; (b) a nucleotide sequence comprising a positive selection marker, such as the marker for neomycine resistance (*neo*); and (c) a second nucleotide sequence that is comprised in the GENSET genomic sequence, and is located on the genome downstream the first GENSET nucleotide sequence (a).

In a preferred embodiment, this DNA construct also comprises a negative selection marker located upstream the nucleotide sequence (a) or downstream the nucleotide sequence (c). Preferably, the negative selection marker comprises the thymidine kinase (*tk*) gene (Thomas *et al.*, 1986), the hygromycine beta gene (Te Riele *et al.*, 1990), the *hprt* gene (Van der Lugt *et al.*, 1991; Reid *et al.*, 1990) or the Diphtheria toxin A fragment (*Dt-A*) gene (Nada *et al.*, 1993; Yagi *et al.*, 1990), which disclosures are hereby incorporated by reference in their entireties. Preferably, the positive selection marker is located within a GENSET exon sequence so as to interrupt the sequence encoding a GENSET protein. These replacement vectors are described, for example, by Thomas *et al.* (1986; 1987), Mansour *et al.* (1988) and Koller *et al.* (1992).

The first and second nucleotide sequences (a) and (c) may be indifferently located within a GENSET regulatory sequence, an intronic sequence, an exon sequence or a sequence containing both regulatory and/or intronic and/or exon sequences. The size of the nucleotide sequences (a) and (c) ranges from 1 to 50 kb, preferably from 1 to 10 kb, more preferably from 2 to 6 kb and most preferably from 2 to 4 kb.

DNA Constructs Allowing Homologous Recombination: Cre-LoxP System.

These new DNA constructs make use of the site specific recombination system of the P1 phage. The P1 phage possesses a recombinase called Cre which interacts specifically with a 34 base pairs *loxP* site. The *loxP* site is composed of two palindromic sequences of 13 bp separated by a 8 bp conserved sequence (Hoess *et al.*, 1986), which disclosure is hereby incorporated by reference in its entirety. The recombination by the Cre enzyme between two *loxP* sites having an identical orientation leads to the deletion of the DNA fragment.

The Cre-*loxP* system used in combination with a homologous recombination technique has been first described by Gu *et al.* (1993, 1994), which disclosures are hereby incorporated by reference in their entireties. Briefly, a nucleotide sequence of interest to be inserted in a targeted location of the genome harbors at least two *loxP* sites in the same orientation and located at the respective ends of a nucleotide sequence to be excised from the recombinant genome. The excision event requires the presence of the recombinase (Cre) enzyme within the nucleus of the recombinant cell host. The recombinase enzyme may be brought at the desired time either by (a) incubating the recombinant cell hosts in a culture medium containing this enzyme, by injecting the Cre enzyme directly into the desired cell, such as described by Araki *et al.* (1995), which disclosure is hereby

incorporated by reference in its entirety, or by lipofection of the enzyme into the cells, such as described by Baubonis *et al.*(1993), which disclosure is hereby incorporated by reference in its entirety; (b) transfecting the cell host with a vector comprising the *Cre* coding sequence operably linked to a promoter functional in the recombinant cell host, which promoter being optionally
5 inducible, said vector being introduced in the recombinant cell host, such as described by Gu *et al.*(1993) and Sauer *et al.*(1988), which disclosures are hereby incorporated by reference in their entireties; (c) introducing in the genome of the cell host a polynucleotide comprising the *Cre* coding sequence operably linked to a promoter functional in the recombinant cell host, which promoter is optionally inducible, and said polynucleotide being inserted in the genome of the cell host either by
10 a random insertion event or an homologous recombination event, such as described by Gu *et al.*(1994).

In a specific embodiment, the vector containing the sequence to be inserted in the GENSET gene by homologous recombination is constructed in such a way that selectable markers are flanked by *loxP* sites of the same orientation, it is possible, by treatment by the *Cre* enzyme, to eliminate the
15 selectable markers while leaving the GENSET sequences of interest that have been inserted by an homologous recombination event. Again, two selectable markers are needed: a positive selection marker to select for the recombination event and a negative selection marker to select for the homologous recombination event. Vectors and methods using the *Cre-loxP* system are described by Zou *et al.*(1994), which disclosure is hereby incorporated by reference in its entirety.

20 Thus, a third preferred DNA construct of the invention comprises, from 5'-end to 3'-end: (a) a first nucleotide sequence that is comprised in the GENSET genomic sequence; (b) a nucleotide sequence comprising a polynucleotide encoding a positive selection marker, said nucleotide sequence comprising additionally two sequences defining a site recognized by a recombinase, such as a *loxP* site, the two sites being placed in the same orientation; and (c) a second nucleotide
25 sequence that is comprised in the GENSET genomic sequence, and is located on the genome downstream of the first GENSET nucleotide sequence (a).

The sequences defining a site recognized by a recombinase, such as a *loxP* site, are preferably located within the nucleotide sequence (b) at suitable locations bordering the nucleotide sequence for which the conditional excision is sought. In one specific embodiment, two *loxP* sites
30 are located at each side of the positive selection marker sequence, in order to allow its excision at a desired time after the occurrence of the homologous recombination event.

In a preferred embodiment of a method using the third DNA construct described above, the excision of the polynucleotide fragment bordered by the two sites recognized by a recombinase, preferably two *loxP* sites, is performed at a desired time, due to the presence within the genome of
35 the recombinant host cell of a sequence encoding the *Cre* enzyme operably linked to a promoter sequence, preferably an inducible promoter, more preferably a tissue-specific promoter sequence

and most preferably a promoter sequence which is both inducible and tissue-specific, such as described by Gu *et al.* (1994).

The presence of the Cre enzyme within the genome of the recombinant cell host may result from the breeding of two transgenic animals, the first transgenic animal bearing the GENSET-derived sequence of interest containing the *loxP* sites as described above and the second transgenic animal bearing the *Cre* coding sequence operably linked to a suitable promoter sequence, such as described by Gu *et al.* (1994).

Spatio-temporal control of the Cre enzyme expression may also be achieved with an adenovirus based vector that contains the Cre gene thus allowing infection of cells, or *in vivo* infection of organs, for delivery of the Cre enzyme, such as described by Anton and Graham (1995) and Kanegae *et al.* (1995), which disclosures are hereby incorporated by reference in their entireties.

The DNA constructs described above may be used to introduce a desired nucleotide sequence of the invention, preferably a GENSET genomic sequence or a GENSET cDNA sequence, and most preferably an altered copy of a GENSET genomic or cDNA sequence, within a predetermined location of the targeted genome, leading either to the generation of an altered copy of a targeted gene (knock-out homologous recombination) or to the replacement of a copy of the targeted gene by another copy sufficiently homologous to allow an homologous recombination event to occur (knock-in homologous recombination).

MODIFYING GENSET EXPRESSION AND/OR BIOLOGICAL ACTIVITY

Modifying endogenous GENSET expression and/or biological activity is expressly contemplated by the present invention.

Screening for compounds that modulate GENSET expression and/or biological activity

The present invention further relates to compounds able to modulate GENSET expression and/or biological activity and methods to use these compounds. Such compounds may interact with the regulatory sequences of GENSET genes or they may interact with GENSET polypeptides directly or indirectly.

Compounds Interacting With GENSET Regulatory Sequences

The present invention also concerns a method for screening substances or molecules that are able to interact with the regulatory sequences of a GENSET gene, such as for example promoter or enhancer sequences in untranscribed regions of the genomic DNA, as determined using any techniques known to those skilled in the art including those described in the section entitled "Identification of Promoters in Cloned Upstream Sequences, or such as regulatory sequences located in untranslated regions of GENSET mRNA.

Sequences within untranscribed or untranslated regions of polynucleotides of the invention may be identified by comparison to databases containing known regulatory sequence such as

transcription start sites, transcription factor binding sites, promoter sequences, enhancer sequences, 5'UTR and 3'UTR elements (Pesole *et al.*, 2000; <http://igs-server.cnrs-mrs.fr/~gauthere/UTR/index.html>). Alternatively, the regulatory sequences of interest may be identified through conventional mutagenesis or deletion analyses of reporter plasmids using, for instance, techniques described in the section entitled "Identification of Promoters in Cloned Upstream Sequences".

Following the identification of potential GENSET regulatory sequences, proteins which interact with these regulatory sequences may be identified as described below.

Gel retardation assays may be performed independently in order to screen candidate molecules that are able to interact with the regulatory sequences of the GENSET gene, such as described by Fried and Crothers (1981), Garner and Revzin (1981) and Dent and Latchman (1993), the teachings of these publications being herein incorporated by reference. These techniques are based on the principle according to which a DNA or mRNA fragment which is bound to a protein migrates slower than the same unbound DNA or mRNA fragment. Briefly, the target nucleotide sequence is labeled. Then the labeled target nucleotide sequence is brought into contact with either a total nuclear extract from cells containing regulation factors, or with different candidate molecules to be tested. The interaction between the target regulatory sequence of the GENSET gene and the candidate molecule or the regulation factor is detected after gel or capillary electrophoresis through a retardation in the migration.

Nucleic acids encoding proteins which are able to interact with the promoter sequence of the GENSET gene, more particularly a nucleotide sequence selected from the group consisting of the polynucleotides of the 5' and 3' regulatory region or a fragment or variant thereof, may be identified by using a one-hybrid system, such as that described in the booklet enclosed in the Matchmaker One-Hybrid System kit from Clontech (Catalog Ref. n° K1603-1), the technical teachings of which are herein incorporated by reference. Briefly, the target nucleotide sequence is cloned upstream of a selectable reporter sequence and the resulting polynucleotide construct is integrated in the yeast genome (*Saccharomyces cerevisiae*). Preferably, multiple copies of the target sequences are inserted into the reporter plasmid in tandem. The yeast cells containing the reporter sequence in their genome are then transformed with a library comprising fusion molecules between cDNAs encoding candidate proteins for binding onto the regulatory sequences of the GENSET gene and sequences encoding the activator domain of a yeast transcription factor such as GAL4. The recombinant yeast cells are plated in a culture broth for selecting cells expressing the reporter sequence. The recombinant yeast cells thus selected contain a fusion protein that is able to bind onto the target regulatory sequence of the GENSET gene. Then, the cDNAs encoding the fusion proteins are sequenced and may be cloned into expression or transcription vectors *in vitro*. The binding of the encoded polypeptides to the target regulatory sequences of the GENSET gene

may be confirmed by techniques familiar to the one skilled in the art, such as gel retardation assays or DNase protection assays.

Ligands interacting with GENSET polypeptides

For the purpose of the present invention, a ligand means a molecule, such as a protein, a peptide, an antibody or any synthetic chemical compound capable of binding to a GENSET protein or one of its fragments or variants or to modulate the expression of the polynucleotide coding for GENSET or a fragment or variant thereof.

In the ligand screening method according to the present invention, a biological sample or a defined molecule to be tested as a putative ligand of a GENSET protein is brought into contact with the corresponding purified GENSET protein, for example the corresponding purified recombinant GENSET protein produced by a recombinant cell host as described herein, in order to form a complex between this protein and the putative ligand molecule to be tested.

As an illustrative example, to study the interaction of a GENSET protein, or a fragment comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of a polypeptide selected from the group consisting of sequences of SEQ ID Nos: 242-482, mature polypeptides included in SEQ ID Nos: 242-272 and 274-384 as well as full-length and mature polypeptides encoded by the clone inserts of the deposited clone pool, with drugs or small molecules, such as molecules generated through combinatorial chemistry approaches, the microdialysis coupled to HPLC method described by Wang *et al.* (1997) or the affinity capillary electrophoresis method described by Bush *et al.* (1997), the disclosures of which are incorporated by reference, can be used.

In further methods, peptides, drugs, fatty acids, lipoproteins, or small molecules which interact with a GENSET protein, or a fragment comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of a polypeptide selected from the group consisting of sequences of SEQ ID Nos: 242-482, mature polypeptides included in SEQ ID Nos: 242-272 and 274-384, as well as full-length and mature polypeptides encoded by the clone inserts of the deposited clone pool may be identified using assays such as the following. The molecule to be tested for binding is labeled with a detectable label, such as a fluorescent, radioactive, or enzymatic tag and placed in contact with immobilized GENSET protein, or a fragment thereof under conditions which permit specific binding to occur. After removal of non-specifically bound molecules, bound molecules are detected using appropriate means.

Various candidate substances or molecules can be assayed for interaction with a GENSET polypeptide. These substances or molecules include, without being limited to, natural or synthetic organic compounds or molecules of biological origin such as polypeptides. When the candidate substance or molecule comprises a polypeptide, this polypeptide may be the resulting expression

product of a phage clone belonging to a phage-based random peptide library, or alternatively the polypeptide may be the resulting expression product of a cDNA library cloned in a vector suitable for performing a two-hybrid screening assay.

A. Candidate ligands obtained from random peptide libraries

- 5 In a particular embodiment of the screening method, the putative ligand is the expression product of a DNA insert contained in a phage vector (Parmley and Smith, 1988). Specifically, random peptide phages libraries are used. The random DNA inserts encode for peptides of 8 to 20 amino acids in length (Oldenburg *et al.*, 1992; Valadon *et al.*, 1996; Lucas, 1994; Westerink, 1995; Felici *et al.*, 1991), which disclosures are hereby incorporated by reference in their entireties.
- 10 According to this particular embodiment, the recombinant phages expressing a protein that binds to an immobilized GENSET protein is retained and the complex formed between the GENSET protein and the recombinant phage may be subsequently immunoprecipitated by a polyclonal or a monoclonal antibody directed against the GENSET protein.

Once the ligand library in recombinant phages has been constructed, the phage population is brought into contact with the immobilized GENSET protein. Then the preparation of complexes is washed in order to remove the non-specifically bound recombinant phages. The phages that bind specifically to the GENSET protein are then eluted by a buffer (acid pH) or immunoprecipitated by the monoclonal antibody produced by the hybridoma anti-GENSET, and this phage population is subsequently amplified by an over-infection of bacteria (for example *E. coli*). The selection step

20 may be repeated several times, preferably 2-4 times, in order to select the more specific recombinant phage clones. The last step comprises characterizing the peptide produced by the selected recombinant phage clones either by expression in infected bacteria and isolation, expressing the phage insert in another host-vector system, or sequencing the insert contained in the selected recombinant phages.

25 *B. Candidate ligands obtained by competition experiments.*

Alternatively, peptides, drugs or small molecules which bind to a GENSET protein or fragment thereof comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of a polypeptide selected from the group consisting of sequences of SEQ ID Nos: 242-482, mature polypeptides

30 included in SEQ ID Nos: 242-272 and 274-384, as well as full-length and mature polypeptides encoded by the clone inserts of the deposited clone pool, may be identified in competition experiments. In such assays, the GENSET protein, or a fragment thereof, is immobilized to a surface, such as a plastic plate. Increasing amounts of the peptides, drugs or small molecules are placed in contact with the immobilized GENSET protein, or a fragment thereof, in the presence of a

35 detectable labeled known GENSET protein ligand. For example, the GENSET ligand may be detectably labeled with a fluorescent, radioactive, or enzymatic tag. The ability of the test molecule

to bind the GENSET protein, or a fragment thereof, is determined by measuring the amount of detectably labeled known ligand bound in the presence of the test molecule. A decrease in the amount of known ligand bound to the GENSET protein, or a fragment thereof, when the test molecule is present indicated that the test molecule is able to bind to the GENSET protein, or a
5 fragment thereof.

C. Candidate ligands obtained by affinity chromatography.

Proteins or other molecules interacting with a GENSET protein, or a fragment thereof comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of a polypeptide selected from the
10 group consisting of sequences of SEQ ID Nos: 242-482, mature polypeptides included in SEQ ID Nos: 242-272 and 274-384, as well as full-length and mature polypeptides encoded by the clone inserts of the deposited clone pool, can also be found using affinity columns which contain the GENSET protein, or a fragment thereof. The GENSET protein, or a fragment thereof, may be attached to the column using conventional techniques including chemical coupling to a suitable
15 column matrix such as agarose, Affi Gel®, or other matrices familiar to those of skill in art. In some embodiments of this method, the affinity column contains chimeric proteins in which the GENSET protein, or a fragment thereof, is fused to glutathion S transferase (GST). A mixture of cellular proteins or pool of expressed proteins as described above is applied to the affinity column. Proteins or other molecules interacting with the GENSET protein, or a fragment thereof, attached to
20 the column can then be isolated and analyzed on 2-D electrophoresis gel as described in Ramunsen *et al.* (1997), the disclosure of which is incorporated by reference. Alternatively, the proteins retained on the affinity column can be purified by electrophoresis based methods and sequenced. The same method can be used to isolate antibodies, to screen phage display products, or to screen phage display human antibodies.

D. Candidate ligands obtained by optical biosensor methods

Proteins interacting with a GENSET protein, or a fragment comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of a polypeptide selected from the group consisting of sequences of SEQ ID Nos: 242-482, mature polypeptides included in SEQ ID Nos: 242-272 and 274-384, as
30 well as full-length and mature polypeptides encoded by the clone inserts of the deposited clone pool, can also be screened by using an Optical Biosensor as described in Edwards and Leatherbarrow (1997) and also in Szabo *et al.* (1995), the disclosures of which are incorporated by reference. This technique permits the detection of interactions between molecules in real time, without the need of labeled molecules. This technique is based on the surface plasmon resonance
35 (SPR) phenomenon. Briefly, the candidate ligand molecule to be tested is attached to a surface (such as a carboxymethyl dextran matrix). A light beam is directed towards the side of the surface

that does not contain the sample to be tested and is reflected by said surface. The SPR phenomenon causes a decrease in the intensity of the reflected light with a specific association of angle and wavelength. The binding of candidate ligand molecules cause a change in the refraction index on the surface, which change is detected as a change in the SPR signal. For screening of candidate
5 ligand molecules or substances that are able to interact with the GENSET protein, or a fragment thereof, the GENSET protein, or a fragment thereof, is immobilized onto a surface. This surface comprises one side of a cell through which flows the candidate molecule to be assayed. The binding of the candidate molecule on the GENSET protein, or a fragment thereof, is detected as a change of the SPR signal. The candidate molecules tested may be proteins, peptides, carbohydrates,
10 lipids, or small molecules generated by combinatorial chemistry. This technique may also be performed by immobilizing eukaryotic or prokaryotic cells or lipid vesicles exhibiting an endogenous or a recombinantly expressed GENSET protein at their surface.

The main advantage of the method is that it allows the determination of the association rate between the GENSET protein and molecules interacting with the GENSET protein. It is thus
15 possible to select specifically ligand molecules interacting with the GENSET protein, or a fragment thereof, through strong or conversely weak association constants.

E. Candidate ligands obtained through a two-hybrid screening assay.

The yeast two-hybrid system is designed to study protein-protein interactions *in vivo* (Fields and Song, 1989), which disclosure is hereby incorporated by reference in its entirety, and relies
20 upon the fusion of a bait protein to the DNA binding domain of the yeast Gal4 protein. This technique is also described in the US Patent N° US 5,667,973 and the US Patent N° 5,283,173, the technical teachings of both patents being herein incorporated by reference.

The general procedure of library screening by the two-hybrid assay may be performed as described by Harper *et al.* (1993) or as described by Cho *et al.* (1998) or also Fromont-Racine *et al.*
25 (1997), which disclosures are hereby incorporated by reference in their entireties.

The bait protein or polypeptide comprises, consists essentially of, or consists of a GENSET polypeptide or a fragment thereof comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of a polypeptide selected from the group consisting of sequences of SEQ ID Nos: 242-482, mature
30 polypeptides included in SEQ ID Nos: 242-272 and 274-384, as well as full-length and mature polypeptides encoded by the clone inserts of the deposited clone pool.

More precisely, the nucleotide sequence encoding the GENSET polypeptide or a fragment or variant thereof is fused to a polynucleotide encoding the DNA binding domain of the GAL4 protein, the fused nucleotide sequence being inserted in a suitable expression vector, for example
35 pAS2 or pM3.

Then, a human cDNA library is constructed in a specially designed vector, such that the human cDNA insert is fused to a nucleotide sequence in the vector that encodes the transcriptional domain of the GAL4 protein. Preferably, the vector used is the pACT vector. The polypeptides encoded by the nucleotide inserts of the human cDNA library are termed "pray" polypeptides.

- 5 A third vector contains a detectable marker gene, such as beta galactosidase gene or CAT gene that is placed under the control of a regulation sequence that is responsive to the binding of a complete Gal4 protein containing both the transcriptional activation domain and the DNA binding domain. For example, the vector pG5EC may be used.

Two different yeast strains are also used. As an illustrative but non limiting example the
10 two different yeast strains may be the followings :

- Y190, the phenotype of which is (MATa, Leu2-3, 112 ura3-12, trp1-901, his3-D200, ade2-101, gal4Dgal180D URA3 GAL-LacZ, LYS GAL-HIS3, cyh^r);
- Y187, the phenotype of which is (MATa gal4 gal80 his3 trp1-901 ade2-101 ura3-52 leu2-3, -112 URA3 GAL-lacZmet^r), which is the opposite mating type of Y190.

- 15 Briefly, 20 µg of pAS2/GENSET and 20 µg of pACT-cDNA library are co-transformed into yeast strain Y190. The transformants are selected for growth on minimal media lacking histidine, leucine and tryptophan, but containing the histidine synthesis inhibitor 3-AT (50 mM). Positive colonies are screened for beta galactosidase by filter lift assay. The double positive colonies (His⁺, beta-gal⁺) are then grown on plates lacking histidine, leucine, but containing
20 tryptophan and cycloheximide (10 mg/ml) to select for loss of pAS2/GENSET plasmids but retention of pACT-cDNA library plasmids. The resulting Y190 strains are mated with Y187 strains expressing GENSET or non-related control proteins; such as cyclophilin B, lamin, or SNF1, as Gal4 fusions as described by Harper *et al.* (1993) and by Bram *et al.* (1993), which disclosures are hereby incorporated by reference in their entireties, and screened for beta galactosidase by filter lift assay.
25 Yeast clones that are beta gal- after mating with the control Gal4 fusions are considered false positives.

- In another embodiment of the two-hybrid method according to the invention, interaction between the GENSET or a fragment or variant thereof with cellular proteins may be assessed using the Matchmaker Two Hybrid System 2 (Catalog No. K1604-1, Clontech). As described in the
30 manual accompanying the kit, the disclosure of which is incorporated herein by reference, nucleic acids encoding the GENSET protein or a portion thereof, are inserted into an expression vector such that they are in frame with DNA encoding the DNA binding domain of the yeast transcriptional activator GAL4. A desired cDNA, preferably human cDNA, is inserted into a second expression vector such that they are in frame with DNA encoding the activation domain of GAL4. The two
35 expression plasmids are transformed into yeast and the yeast are plated on selection medium which selects for expression of selectable markers on each of the expression vectors as well as GAL4 dependent expression of the HIS3 gene. Transformants capable of growing on medium lacking

histidine are screened for GAL4 dependent lacZ expression. Those cells which are positive in both the histidine selection and the lacZ assay contain interaction between GENSET and the protein or peptide encoded by the initially selected cDNA insert.

Compounds Modulating GENSET biological activity

5 Another method of screening for compounds that modulate GENSET gene expression and/or biological activity is by measuring the effects of test compounds on a given cellular property in a host cell, such as apoptosis, proliferation, differentiation, protein glycosylation, etc... using a variety of techniques known to those skilled in the art including those described herein and especially in the section entitled “**Erreur! Source du renvoi introuvable.**”.

10 In one embodiment, the present invention relates to a method of identifying an agent which alters GENSET activity, wherein a nucleic acid construct comprising a nucleic acid which encodes a mammalian GENSET polypeptide is introduced into a host cell. The host cells produced are maintained under conditions appropriate for expression of the encoded mammalian GENSET polypeptides, whereby the nucleic acid is expressed. The host cells are then contacted with a
15 compound to be assessed (an agent) and the given cellular property of the cells is detected in the presence of the compound to be assessed. Detection of a change in the given cellular property in the presence of the agent indicates that the agent alters GENSET activity.

In a particular embodiment, the invention relates to a method of identifying an agent which is an activator of GENSET activity, wherein detection of a change of the given cellular property in
20 the presence of the agent indicates that the agent activates GENSET activity. In another particular embodiment, the invention relates to a method of identifying an agent which is an inhibitor of GENSET activity, wherein detection of a change of the given cellular property in the presence of the agent indicates that the agent inhibits GENSET activity.

Methods of Screening for Compounds Modulating GENSET Expression and/or Activity

25 The present invention also relates to methods of screening compounds for their ability to modulate (e.g. increase or inhibit) the activity or expression of GENSET. More specifically, the present invention relates to methods of testing compounds for their ability either to increase or to decrease expression or activity of GENSET. The assays are performed *in vitro* or *in vivo*.

In vitro methods

30 *In vitro*, cells expressing GENSET are incubated in the presence and absence of the test compound. By determining the level of GENSET expression in the presence of the test compound or the level of GENSET activity in the presence of the test compound, compounds can be identified that suppress or enhance GENSET expression or activity. Alternatively, constructs comprising a GENSET regulatory sequence operably linked to a reporter gene (e.g. luciferase, chloramphenicol
35 acetyl transferase, LacZ, green fluorescent protein, etc.) can be introduced into host cells and the

effect of the test compounds on expression of the reporter gene detected. Cells suitable for use in the foregoing assays include, but are not limited to, cells having the same origin as tissues or cell lines in which the polypeptide is known to be expressed using the data from Table IX.

Consequently, the present invention encompasses a method for screening molecules that
5 modulate the expression of a GENSET gene, said screening method comprising the steps of:

- a) cultivating a prokaryotic or an eukaryotic cell that has been transfected with a nucleotide sequence encoding a GENSET protein or a variant or a fragment thereof, placed under the control of its own promoter;
- b) bringing into contact said cultivated cell with a molecule to be tested;
- 10 c) quantifying the expression of said GENSET protein or a variant or a fragment thereof in the presence of said molecule.

Using DNA recombination techniques well known by the one skilled in the art, the GENSET protein encoding DNA sequence is inserted into an expression vector, downstream from its promoter sequence. As an illustrative example, the promoter sequence of the GENSET gene is
15 contained in the 5' untranscribed region of the GENSET genomic DNA.

The quantification of the expression of a GENSET protein may be realized either at the mRNA level (using for example Northern blots, RT-PCR, preferably quantitative RT-PCR with primers and probes specific for the GENSET mRNA of interest) or at the protein level (using polyclonal or monoclonal antibodies in immunoassays such as ELISA or RIA assays, Western blots,
20 or immunochemistry).

The present invention also concerns a method for screening substances or molecules that are able to increase, or in contrast to decrease, the level of expression of a GENSET gene. Such a method may allow the one skilled in the art to select substances exerting a regulating effect on the expression level of a GENSET gene and which may be useful as active ingredients included in
25 pharmaceutical compositions for treating patients suffering from disorders associated with abnormal levels of GENSET products.

Thus, also part of the present invention is a method for screening a candidate molecule that modulates the expression of a GENSET gene, this method comprises the following steps:

- a) providing a recombinant cell host containing a nucleic acid, wherein said nucleic acid
30 comprises a GENSET 5' regulatory region or a regulatory active fragment or variant thereof, operably linked to a polynucleotide encoding a detectable protein;
- b) obtaining a candidate molecule; and
- c) determining the ability of said candidate molecule to modulate the expression levels of said polynucleotide encoding the detectable protein.

35 In a further embodiment, said nucleic acid comprising a GENSET 5' regulatory region or a regulatory active fragment or variant thereof, includes the 5'UTR region of a GENSET cDNA selected from the group comprising of the 5'UTRs of the sequences of SEQ ID Nos 1-241,

sequences of clones inserts of the deposited clone pool, regulatory active fragments and variants thereof. In a more preferred embodiment of the above screening method, said nucleic acid includes a promoter sequence which is endogenous with respect to the GENSET 5'UTR sequence. In another more preferred embodiment of the above screening method, said nucleic acid includes a
5 promoter sequence which is exogenous with respect to the GENSET 5'UTR sequence defined therein.

Preferred polynucleotides encoding a detectable protein are polynucleotides encoding beta galactosidase, green fluorescent protein (GFP) and chloramphenicol acetyl transferase (CAT).

The invention further relates to a method for the production of a pharmaceutical
10 composition comprising a method of screening a candidate molecule that modulates the expression of a GENSET gene and furthermore mixing the identified molecule with a pharmaceutically acceptable carrier.

The invention also pertains to kits for the screening of a candidate substance modulating the expression of a GENSET gene. Preferably, such kits comprise a recombinant vector that allows the
15 expression of a GENSET 5' regulatory region or a regulatory active fragment or a variant thereof, operably linked to a polynucleotide encoding a detectable protein or a GENSET protein or a fragment or a variant thereof. More preferably, such kits include a recombinant vector that comprises a nucleic acid including the 5'UTR region of a GENSET cDNA selected from the group comprising the 5'UTRs of the sequences of SEQ ID Nos 1-241, sequences of clones inserts of the
20 deposited clone pool, regulatory active fragments and variants thereof, being operably linked to a polynucleotide encoding a detectable protein.

For the design of suitable recombinant vectors useful for performing the screening methods described above, it will be referred to the section of the present specification wherein the preferred recombinant vectors of the invention are detailed.

25 Another object of the present invention comprises methods and kits for the screening of candidate substances that interact with a GENSET polypeptide, fragments or variants thereof. By their capacity to bind covalently or non-covalently to a GENSET protein, fragments or variants thereof, these substances or molecules may be advantageously used both *in vitro* and *in vivo*.

In vitro, said interacting molecules may be used as detection means in order to identify the
30 presence of a GENSET protein in a sample, preferably a biological sample.

A method for the screening of a candidate substance that interact with a GENSET polypeptide, fragments or variants thereof, said methods comprising the following steps:

a) providing a polypeptide comprising, consisting essentially of, or consisting of a GENSET protein or a fragment comprising a contiguous span of at least 6 amino acids, preferably at least 8 to
35 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of a polypeptide selected from the group consisting of sequences of SEQ ID Nos: 242-482, mature polypeptides

included in SEQ ID Nos: 242-272 and 274-384 as well as full-length and mature polypeptides encoded by the clone inserts of the deposited clone pool;

- b) obtaining a candidate substance;
- c) bringing into contact said polypeptide with said candidate substance;
- 5 d) detecting the complexes formed between said polypeptide and said candidate substance.

The invention further relates to a method for the production of a pharmaceutical composition comprising a method for the screening of a candidate substance that interact with a GENSET polypeptide, fragments or variants thereof and furthermore mixing the identified substance with a pharmaceutically acceptable carrier.

- 10 The invention further concerns a kit for the screening of a candidate substance interacting with the GENSET polypeptide, wherein said kit comprises:

- a) a polypeptide comprising, consisting essentially of, or consisting of a GENSET protein or a fragment comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of a polypeptide
- 15 selected from the group consisting of sequences of SEQ ID Nos: 242-482, mature polypeptides included in SEQ ID Nos: 242-272 and 274-384 as well as full-length and mature polypeptides encoded by the clone inserts of the deposited clone pool; and
- b) optionally means useful to detect the complex formed between said polypeptide or a variant thereof and the candidate substance.

- 20 In a preferred embodiment of the kit described above, the detection means comprises a monoclonal or polyclonal antibody binding to said GENSET protein or fragment or variant thereof.

In vivo methods

- Compounds that suppress or enhance GENSET expression can also be identified using *in vivo* screens. In these assays, the test compound is administered (e.g. IV, IP, IM, orally, or
- 25 otherwise), to the animal, for example, at a variety of dose levels. The effect of the compound on GENSET expression is determined by comparing GENSET levels, for example in tissues known to express the gene of interest using, for example the data obtained in Table IX, and using Northern blots, immunoassays, PCR, etc., as described above. Suitable test animals include rodents (e.g., mice and rats), primates, mammals. Humanized mice can also be used as test animals, that is mice
- 30 in which the endogenous mouse protein is ablated (knocked out) and the homologous human protein added back by standard transgenic approaches. Such mice express only the human form of a protein. Humanized mice expressing only the human GENSET can be used to study *in vivo* responses to potential agents regulating GENSET protein or mRNA levels. As an example, transgenic mice have been produced carrying the human apoE4 gene. They are then bred with a
- 35 mouse line that lacks endogenous apoE, to produce an animal model carrying human proteins believed to be instrumental in development of Alzheimer's pathology. Such transgenic animals are useful for dissecting the biochemical and physiological steps of disease, and for development of

therapies for disease intervention (Loring, *et al*, 1996) (incorporated herein by reference in its entirety).

Uses for compounds modulating GENSET expression and/or biological activity

Using *in vivo* (or *in vitro*) systems, it may be possible to identify compounds that exert a
5 tissue specific effect, for example, that increase GENSET expression or activity only in tissues of
interest. Screening procedures such as those described above are also useful for identifying agents
for their potential use in pharmacological intervention strategies. Agents that enhance GENSET
expression or stimulate its activity may thus be used to treat disorders which require upregulated
levels of GENSET gene expression and/or activity. Compounds that suppress GENSET expression
10 or inhibit its activity can be used to treat disorders which require downregulated levels of GENSET
gene expression and/or activity.

Also encompassed by the present invention is an agent which interacts with GENSET
directly or indirectly, and inhibits or enhances GENSET expression and/or function. In one
embodiment, the agent is an inhibitor which interferes with GENSET directly (e.g., by binding
15 GENSET) or indirectly (e.g., by blocking the ability of GENSET to have a GENSET biological
activity). In a particular embodiment, an inhibitor of GENSET protein is an antibody specific for
GENSET protein or a functional portion of GENSET; that is, the antibody binds a GENSET
polypeptide. For example, the antibody can be specific for a polypeptide encoded by one of the
amino acid sequences of human GENSET genes (SEQ ID Nos: 242-482, mature polypeptides
20 included in SEQ ID Nos: 242-272 and 274-384, full-length and mature polypeptides encoded by the
clone inserts of the deposited clone pool), mammal GENSET or portions thereof. Alternatively, the
inhibitor can be an agent other than an antibody (e.g., small organic molecule, protein or peptide)
which binds GENSET and blocks its activity. For example, the inhibitor can be an agent which
mimics GENSET structurally, but lacks its function. Alternatively, it can be an agent which binds
25 to or interacts with a molecule which GENSET normally binds with or interacts with, thus blocking
GENSET from doing so and preventing it from exerting the effects it would normally exert.

In another embodiment, the agent is an enhancer (activator) of GENSET which increases
the activity of GENSET (increases the effect of a given amount or level of GENSET), increases the
length of time it is effective (by preventing its degradation or otherwise prolonging the time during
30 which it is active) or both either directly or indirectly.

The GENSET sequences of the present invention can also be used to generate nonhuman
gene knockout animals, such as mice, which lack a GENSET gene or transgenically overexpress
GENSET. For example, such GENSET gene knockout mice can be generated and used to obtain
further insight into the function of GENSET as well as assess the specificity of GENSET activators
35 and inhibitors. Also, over expression of GENSET (e.g., human GENSET) in transgenic mice can
be used as a means of creating a test system for GENSET activators and inhibitors (e.g., against

human GENSET). In addition, the GENSET gene can be used to clone the GENSET promoter/enhancer in order to identify regulators of GENSET transcription. GENSET gene knockout animals include animals which completely or partially lack the GENSET gene and/or GENSET activity or function. Thus the present invention relates to a method of inhibiting (partially or completely) GENSET biological activity in a mammal (e.g., human) comprising administering to the mammal an effective amount of an inhibitor of GENSET. The invention also relates to a method of enhancing GENSET biological activity in a mammal comprising administering to the mammal an effective amount of an enhancer GENSET.

Inhibiting GENSET expression

10 Therapeutic compositions according to the present invention may comprise advantageously one or several GENSET oligonucleotide fragments as an antisense tool or a triple helix tool that inhibits the expression of the corresponding GENSET gene.

Antisense Approach

15 In antisense approaches, nucleic acid sequences complementary to an mRNA are hybridized to the mRNA intracellularly, thereby blocking the expression of the protein encoded by the mRNA. The antisense nucleic acid molecules to be used in gene therapy may be either DNA or RNA sequences. Preferred methods using antisense polynucleotide according to the present invention are the procedures described by Sczakiel *et al.*(1995), which disclosure is hereby incorporated by reference in its entirety.

20 Preferably, the antisense tools are chosen among the polynucleotides (15-200 bp long) that are complementary to GENSET mRNA, more preferably to the 5' end of the GENSET mRNA. In another embodiment, a combination of different antisense polynucleotides complementary to different parts of the desired targeted gene are used.

25 Other preferred antisense polynucleotides according to the present invention are sequences complementary to either a sequence of GENSET mRNAs comprising the translation initiation codon ATG or a sequence of GENSET genomic DNA containing a splicing donor or acceptor site.

30 Preferably, the antisense polynucleotides of the invention have a 3' polyadenylation signal that has been replaced with a self-cleaving ribozyme sequence, such that RNA polymerase II transcripts are produced without poly(A) at their 3' ends, these antisense polynucleotides being incapable of export from the nucleus, such as described by Liu *et al.*(1994), which disclosure is hereby incorporated by reference in its entirety. In a preferred embodiment, these GENSET antisense polynucleotides also comprise, within the ribozyme cassette, a histone stem-loop structure to stabilize cleaved transcripts against 3'-5' exonucleolytic degradation, such as the structure described by Eckner *et al.*(1991), which disclosure is hereby incorporated by reference in its
35 entirety.

The antisense nucleic acids should have a length and melting temperature sufficient to permit formation of an intracellular duplex having sufficient stability to inhibit the expression of the GENSET mRNA in the duplex. Strategies for designing antisense nucleic acids suitable for use in gene therapy are disclosed in Green *et al.*, (1986) and Izant and Weintraub, (1984), the disclosures of which are incorporated herein by reference.

In some strategies, antisense molecules are obtained by reversing the orientation of the GENSET coding region with respect to a promoter so as to transcribe the opposite strand from that which is normally transcribed in the cell. The antisense molecules may be transcribed using *in vitro* transcription systems such as those which employ T7 or SP6 polymerase to generate the transcript.

10 Another approach involves transcription of GENSET antisense nucleic acids *in vivo* by operably linking DNA containing the antisense sequence to a promoter in a suitable expression vector.

Alternatively, oligonucleotides which are complementary to the strand normally transcribed in the cell may be synthesized *in vitro*. Thus, the antisense nucleic acids are complementary to the corresponding mRNA and are capable of hybridizing to the mRNA to create a duplex. In some embodiments, the antisense sequences may contain modified sugar phosphate backbones to increase stability and make them less sensitive to RNase activity. Examples of modifications suitable for use in antisense strategies include 2' O-methyl RNA oligonucleotides and Protein-nucleic acid (PNA) oligonucleotides. Further examples are described by Rossi *et al.*, (1991), which disclosure is hereby incorporated by reference in its entirety.

20 Various types of antisense oligonucleotides complementary to the sequence of the GENSET cDNA or genomic DNA may be used. In one preferred embodiment, stable and semi-stable antisense oligonucleotides described in International Application No. PCT WO94/23026, hereby incorporated by reference, are used. In these molecules, the 3' end or both the 3' and 5' ends are engaged in intramolecular hydrogen bonding between complementary base pairs. These molecules are better able to withstand exonuclease attacks and exhibit increased stability compared to conventional antisense oligonucleotides.

In another preferred embodiment, the antisense oligodeoxynucleotides against herpes simplex virus types 1 and 2 described in International Application No. WO 95/04141, hereby incorporated by reference, are used.

30 In yet another preferred embodiment, the covalently cross-linked antisense oligonucleotides described in International Application No. WO 96/31523, hereby incorporated by reference, are used. These double- or single-stranded oligonucleotides comprise one or more, respectively, inter- or intra-oligonucleotide covalent cross-linkages, wherein the linkage consists of an amide bond between a primary amine group of one strand and a carboxyl group of the other strand or of the same strand, respectively, the primary amine group being directly substituted in the 2' position of the strand nucleotide monosaccharide ring, and the carboxyl group being carried by an aliphatic

spacer group substituted on a nucleotide or nucleotide analog of the other strand or the same strand, respectively.

The antisense oligodeoxynucleotides and oligonucleotides disclosed in International Application No. WO 92/18522, incorporated by reference, may also be used. These molecules are stable to degradation and contain at least one transcription control recognition sequence which binds to control proteins and are effective as decoys therefor. These molecules may contain "hairpin" structures, "dumbbell" structures, "modified dumbbell" structures, "cross-linked" decoy structures and "loop" structures.

In another preferred embodiment, the cyclic double-stranded oligonucleotides described in European Patent Application No. 0 572 287 A2, hereby incorporated by reference are used. These ligated oligonucleotide "dumbbells" contain the binding site for a transcription factor and inhibit expression of the gene under control of the transcription factor by sequestering the factor.

Use of the closed antisense oligonucleotides disclosed in International Application No. WO 92/19732, hereby incorporated by reference, is also contemplated. Because these molecules have no free ends, they are more resistant to degradation by exonucleases than are conventional oligonucleotides. These oligonucleotides may be multifunctional, interacting with several regions which are not adjacent to the target mRNA.

The appropriate level of antisense nucleic acids required to inhibit gene expression may be determined using *in vitro* expression analysis. The antisense molecule may be introduced into the cells by diffusion, injection, infection or transfection using procedures known in the art. For example, the antisense nucleic acids can be introduced into the body as a bare or naked oligonucleotide, oligonucleotide encapsulated in lipid, oligonucleotide sequence encapsidated by viral protein, or as an oligonucleotide operably linked to a promoter contained in an expression vector. The expression vector may be any of a variety of expression vectors known in the art, including retroviral or viral vectors, vectors capable of extrachromosomal replication, or integrating vectors. The vectors may be DNA or RNA.

The antisense molecules are introduced onto cell samples at a number of different concentrations preferably between $1 \times 10^{-10} \text{M}$ to $1 \times 10^{-4} \text{M}$. Once the minimum concentration that can adequately control gene expression is identified, the optimized dose is translated into a dosage suitable for use *in vivo*. For example, an inhibiting concentration in culture of 1×10^{-7} translates into a dose of approximately 0.6 mg/kg bodyweight. Levels of oligonucleotide approaching 100 mg/kg bodyweight or higher may be possible after testing the toxicity of the oligonucleotide in laboratory animals. It is additionally contemplated that cells from the vertebrate are removed, treated with the antisense oligonucleotide, and reintroduced into the vertebrate.

In a preferred application of this invention, the polypeptide encoded by the gene is first identified, so that the effectiveness of antisense inhibition on translation can be monitored using

techniques that include but are not limited to antibody-mediated tests such as RIAs and ELISA, functional assays, or radiolabeling.

An alternative to the antisense technology that is used according to the present invention comprises using ribozymes that will bind to a target sequence via their complementary polynucleotide tail and that will cleave the corresponding RNA by hydrolyzing its target site (namely "hammerhead ribozymes"). Briefly, the simplified cycle of a hammerhead ribozyme comprises (1) sequence specific binding to the target RNA via complementary antisense sequences; (2) site-specific hydrolysis of the cleavable motif of the target strand; and (3) release of cleavage products, which gives rise to another catalytic cycle. Indeed, the use of long-chain antisense polynucleotide (at least 30 bases long) or ribozymes with long antisense arms are advantageous. A preferred delivery system for antisense ribozyme is achieved by covalently linking these antisense ribozymes to lipophilic groups or to use liposomes as a convenient vector. Preferred antisense ribozymes according to the present invention are prepared as described by Rossi *et al.*, (1991) and Sczakiel *et al.* (1995), the specific preparation procedures being referred to in said articles being herein incorporated by reference.

Triple Helix Approach

The GENSET genomic DNA may also be used to inhibit the expression of the GENSET gene based on intracellular triple helix formation.

Triple helix oligonucleotides are used to inhibit transcription from a genome. They are particularly useful for studying alterations in cell activity when it is associated with a particular gene. The GENSET cDNAs or genomic DNAs of the present invention or, more preferably, a fragment of those sequences, can be used to inhibit gene expression in individuals having diseases associated with expression of a particular gene. Similarly, a portion of the GENSET genomic DNA can be used to study the effect of inhibiting GENSET transcription within a cell. Traditionally, homopurine sequences were considered the most useful for triple helix strategies. However, homopyrimidine sequences can also inhibit gene expression. Such homopyrimidine oligonucleotides bind to the major groove at homopurine:homopyrimidine sequences. Thus, both types of sequences from the GENSET genomic DNA are contemplated within the scope of this invention.

To carry out gene therapy strategies using the triple helix approach, the sequences of the GENSET genomic DNA are first scanned to identify 10-mer to 20-mer homopyrimidine or homopurine stretches which could be used in triple-helix based strategies for inhibiting GENSET expression. Following identification of candidate homopyrimidine or homopurine stretches, their efficiency in inhibiting GENSET expression is assessed by introducing varying amounts of oligonucleotides containing the candidate sequences into tissue culture cells which express the GENSET gene.

The oligonucleotides can be introduced into the cells using a variety of methods known to those skilled in the art, including but not limited to calcium phosphate precipitation, DEAE-Dextran, electroporation, liposome-mediated transfection or native uptake.

Treated cells are monitored for altered cell function or reduced GENSET expression using techniques such as Northern blotting, RNase protection assays, or PCR based strategies to monitor the transcription levels of the GENSET gene in cells which have been treated with the oligonucleotide. The cell functions to be monitored are predicted based upon the homologies of the target gene corresponding to the cDNA from which the oligonucleotide was derived with known gene sequences that have been associated with a particular function. The cell functions can also be predicted based on the presence of abnormal physiology within cells derived from individuals with a particular inherited disease, particularly when the cDNA is associated with the disease using techniques described in the section entitled "Identification of genes associated with hereditary diseases or drug response".

The oligonucleotides which are effective in inhibiting gene expression in tissue culture cells may then be introduced *in vivo* using the techniques and at a dosage calculated based on the *in vitro* results, as described in the section entitled "Antisense Approach".

In some embodiments, the natural (beta) anomers of the oligonucleotide units can be replaced with alpha anomers to render the oligonucleotide more resistant to nucleases. Further, an intercalating agent such as ethidium bromide, or the like, can be attached to the 3' end of the alpha oligonucleotide to stabilize the triple helix. For information on the generation of oligonucleotides suitable for triple helix formation see Griffin *et al.* (1989), which is hereby incorporated by this reference.

Treating GENSET-related disorders

The present invention further relates to methods of treating diseases/disorders by increasing GENSET activity and/or expression. The invention also relates to methods of treating diseases/disorders by decreasing GENSET activity and or expression. These methodologies can be effected using compounds selected using screening protocols such as those described herein and/or by using the gene therapy and antisense approaches described in the art and herein. Gene therapy can be used to effect targeted expression of GENSET. The GENSET coding sequence can be cloned into an appropriate expression vector and targeted to a particular cell type(s) to achieve efficient, high level expression. Introduction of the GENSET coding sequence into target cells can be achieved, for example, using particle mediated DNA delivery, (Haynes, 1996 and Maurer, 1999), direct injection of naked DNA, (Levy *et al.*, 1996; and Felgner, 1996), or viral vector mediated transport (Smith *et al.*, 1996, Stone *et al.*, 2000; Wu and Atai, 2000), each of which disclosures are hereby incorporated by reference in their entireties. Tissue specific effects can be achieved, for

example, in the case of virus mediated transport by using viral vectors that are tissue specific, or by the use of promoters that are tissue specific.

Combinatorial approaches can also be used to ensure that the GENSET coding sequence is activated in the target tissue (Butt and Karathanasis, 1995; Miller and Whelan, 1997), which
5 disclosures are hereby incorporated by reference in their entireties. Antisense oligonucleotides complementary to GENSET mRNA can be used to selectively diminish or ablate the expression of the protein, for example, at sites of inflammation. More specifically, antisense constructs or antisense oligonucleotides can be used to inhibit the production of GENSET in high expressing cells such as those cited in the third column of Table X. Antisense mRNA can be produced by
10 transfecting into target cells an expression vector with the GENSET gene sequence, or portion thereof, oriented in an antisense direction relative to the direction of transcription. Appropriate vectors include viral vectors, including retroviral, adenoviral, and adeno-associated viral vectors, as well as nonviral vectors. Tissue specific promoters can be used. Alternatively, antisense oligonucleotides can be introduced directly into target cells to achieve the same goal. (See also
15 other delivery methodologies described herein in connection with gene therapy.). Oligonucleotides can be selected/designed to achieve a high level of specificity (Wagner *et al.*, 1996), which disclosure is hereby incorporated by reference in its entirety. The therapeutic methodologies described herein are applicable to both human and non-human mammals (including cats and dogs).

PHARMACEUTICAL AND PHYSIOLOGICALLY ACCEPTABLE COMPOSITIONS

20 The present invention also relates to pharmaceutical or physiologically acceptable compositions comprising, as active agent, the polypeptides, nucleic acids or antibodies of the invention. The invention also relates to compositions comprising, as active agent, compounds selected using the above-described screening protocols. Such compositions include the active agent in combination with a pharmaceutical or physiologically acceptably acceptable carrier. In the case
25 of naked DNA, the "carrier" may be gold particles. The amount of active agent in the composition can vary with the agent, the patient and the effect sought. Likewise, the dosing regimen can vary depending on the composition and the disease/disorder to be treated.

Therefore, the invention related to methods for the production of pharmaceutical composition comprising a method for selecting an active agent, compound, substance or molecule
30 using any of the screening method described herein and furthermore mixing the identified active agent, compound, substance or molecule with a pharmaceutically acceptable carrier.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal,
35 enteral, topical, sublingual, or rectal means. In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically acceptable carriers comprising

excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack PublishingCo. Easton, Pa).

5 Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

10 Pharmaceutical preparations for oral use can be obtained through a combination of active compounds with solid excipient, sulting mixture is optionally grinding, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, 15 hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums including arabic and tragacanth; and proteins such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate.

Dragee cores may be used in conjunction with suitable coatings, such as concentrated sugar 20 solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titaniumdioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

Pharmaceutical preparations which can be used orally include push-fit capsules made of 25 gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquidpolyethylene glycol with or without stabilizers.

30 Pharmaceutical formulations suitable for parenteral administration may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethylcellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as 35 appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes.

Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

5 The pharmaceutical compositions of the present invention may be manufactured in a manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes.

 The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, succinic, etc.

10 Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preferred preparation may be a lyophilized powder which may contain any or all of the following: 1-50 mM histidine, 0.1%-2% sucrose, and 2-7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

 After pharmaceutical compositions have been prepared, they can be placed in an

15 appropriate container and labeled for treatment of an indicated condition. For administration of GENSET, such labeling would include amount, frequency, and method of administration.

 Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

20 For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models, usually mice, rabbits, dogs, or pigs. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

25 A therapeutically effective dose refers to that amount of active ingredient, for example GENSET or fragments thereof, antibodies of GENSET, agonists, antagonists or inhibitors of GENSET, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to

30 50% of the population). The dose ratio between therapeutic and toxic effects is the therapeutic index, and it can be expressed as the ratio, LD50/ED50. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies is used in formulating a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that include the ED50

35 with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject that requires treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, general health of the subject, age, weight, and gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

15 USES OF GENSET SEQUENCES: COMPUTER-RELATED EMBODIMENTS

As used herein the term "cDNA codes of SEQ ID Nos: 1-241" encompasses the nucleotide sequences of SEQ ID Nos: 1-241 and of clones inserts of the deposited clone pool, fragments thereof, nucleotide sequences homologous thereto, and sequences complementary to all of the preceding sequences. The fragments include fragments of SEQ ID Nos: 1-241 comprising at least 8, 10, 12, 15, 18, 20, 25, 28, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, 500, 1000 or 2000 consecutive nucleotides of SEQ ID Nos: 1-241. Preferably the fragments include signal sequences and coding sequences for mature polypeptides of SEQ ID Nos: 1-31 and 33-143, polynucleotides described in Tables Va and Table Vb, polynucleotides encoding polypeptides described in Table VI, polynucleotide described herein as encoding polypeptides having a biological activity, or fragments comprising at least 8, 10, 12, 15, 18, 20, 25, 28, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, 500, 1000 or 2000 consecutive nucleotides of the signal sequences or coding sequences for mature polypeptides of SEQ ID Nos: 1-31 and 33-143, polynucleotides described in Tables Va and Table Vb, polynucleotides encoding polypeptides described in Table VI, and polynucleotide described herein as encoding polypeptides having a biological activity. Homologous sequences and fragments of SEQ ID Nos: 1-241 refer to a sequence having at least 99%, 98%, 97%, 96%, 95%, 90%, 85%, 80%, or 75% identity to these sequences. Identity may be determined using any of the computer programs and parameters described herein, including BLAST2N with the default parameters or with any modified parameters. Homologous sequences also include RNA sequences in which uridines replace the thymines in the cDNA codes of SEQ ID Nos: 1-241. The homologous sequences may be obtained using any of the procedures described herein or may result from the correction of a sequencing error as described above. Preferably the homologous sequences and fragments of SEQ

ID Nos: 1-241 include polynucleotides homologous to signal sequences and coding sequences for mature polypeptides of SEQ ID Nos: 1-31 and 33-143, polynucleotides described in Tables Va and Table Vb, polynucleotides encoding a polypeptide fragment described as a domain in Table VI, polynucleotide described herein as encoding polypeptides having a biological activity, or fragments comprising at least 8, 10, 12, 15, 18, 20, 25, 28, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, 500, 1000 or 2000 consecutive nucleotides of the signal sequences and coding sequences for mature polypeptides of SEQ ID Nos: 1-31 and 33-143, polynucleotides described in Tables Va and Table Vb, polynucleotides described in Table VI, and polynucleotide described herein as encoding polypeptides having a biological activity. It will be appreciated that the cDNA codes of SEQ ID Nos: 1-241 can be represented in the traditional single character format (See the inside back cover of Styer, 1995) or in any other format which records the identity of the nucleotides in a sequence.

As used herein the term “polypeptide codes of SEQ ID Nos: 242-482” encompasses the polypeptide sequences of SEQ ID Nos: 242-482, the signal peptides included in SEQ ID Nos: 242-272 and 274-384, the mature polypeptides included in SEQ ID Nos: 242-272 and 274-384, the full-length, signal peptides and mature polypeptide sequences encoded by the clone inserts of the deposited clone pool, polypeptide sequences homologous thereto, or fragments of any of the preceding sequences. Homologous polypeptide sequences refer to a polypeptide sequence having at least 99%, 98%, 97%, 96%, 95%, 90%, 85%, 80%, 75% identity to one of the polypeptide sequences of SEQ ID Nos: 242-482, the signal peptides included in SEQ ID Nos: 242-272 and 274-384, the mature polypeptides included in SEQ ID Nos: 242-272 and 274-384, the full-length, signal peptides and mature polypeptide sequences encoded by the clone inserts of the deposited clone pool. Identity may be determined using any of the computer programs and parameters described herein, including FASTA with the default parameters or with any modified parameters. The homologous sequences may be obtained using any of the procedures described herein or may result from the correction of a sequencing error as described above. The polypeptide fragments comprise at least 5, 6, 8, 10, 12, 15, 20, 25, 30, 35, 40, 50, 60, 75, 100, 150, 200, 250, 300, 350, 400, 450 or 500 consecutive amino acids of the polypeptides of SEQ ID Nos: 242-482. Preferably, the fragments include polypeptides encoded by the signal peptides included in SEQ ID Nos: 242-272 and 274-384, mature polypeptides included in SEQ ID Nos: 242-272 and 274-384, polynucleotides described in Tables Va and in Table Vb, domains described in Table VI, epitopes described in Table VII, polypeptides described herein as having a biological activity, or fragments comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 150, 200, 300 or 400 consecutive amino acids of the signal peptides included in SEQ ID Nos: 242-272 and 274-384, mature polypeptides included in SEQ ID Nos: 242-272 and 274-384, the polypeptides encoded by the polynucleotides described in Tables Va and in Table Vb, domains of Table VI, epitopes of Table VII or of polypeptides described herein as having a biological activity. It will be appreciated that the polypeptide codes of the SEQ ID Nos: 242-482 can be represented in the traditional single character format or three letter

format (See the inside back cover of Stryer, 1995) or in any other format which relates the identity of the polypeptides in a sequence.

It will be appreciated by those skilled in the art that the nucleic acid codes of the invention and polypeptide codes of the invention can be stored, recorded, and manipulated on any medium which can be read and accessed by a computer. As used herein, the words "recorded" and "stored" refer to a process for storing information on a computer medium. A skilled artisan can readily adopt any of the presently known methods for recording information on a computer readable medium to generate manufactures comprising one or more of the nucleic acid codes of the invention, or one or more of the polypeptide codes of the invention. Another aspect of the present invention is a computer readable medium having recorded thereon at least 2, 5, 10, 15, 20, 25, 30, 50, 75, 100, 150 or 200 nucleic acid codes of the invention. Another aspect of the present invention is a computer readable medium having recorded thereon at least 2, 5, 10, 15, 20, 25, 30, 50, 75, 100, 150 or 200 polypeptide codes of the invention.

Computer readable media include magnetically readable media, optically readable media, electronically readable media and magnetic/optical media. For example, the computer readable media may be a hard disk, a floppy disk, a magnetic tape, CD-ROM, Digital Versatile Disk (DVD), Random Access Memory (RAM), or Read Only Memory (ROM) as well as other types of other media known to those skilled in the art.

Embodiments of the present invention include systems, particularly computer systems which store and manipulate the sequence information described herein. One example of a computer system 100 is illustrated in block diagram form in Figure 2. As used herein, "a computer system" refers to the hardware components, software components, and data storage components used to analyze the nucleotide sequences of the nucleic acid codes of the invention or the amino acid sequences of the polypeptide codes of the invention. In one embodiment, the computer system 100 is a Sun Enterprise 1000 server (Sun Microsystems, Palo Alto, CA). The computer system 100 preferably includes a processor for processing, accessing and manipulating the sequence data. The processor 105 can be any well-known type of central processing unit, such as the Pentium III from Intel Corporation, or similar processor from Sun, Motorola, Compaq or International Business Machines.

Preferably, the computer system 100 is a general purpose system that comprises the processor 105 and one or more internal data storage components 110 for storing data, and one or more data retrieving devices for retrieving the data stored on the data storage components. A skilled artisan can readily appreciate that any one of the currently available computer systems are suitable.

In one particular embodiment, the computer system 100 includes a processor 105 connected to a bus which is connected to a main memory 115 (preferably implemented as RAM) and one or more internal data storage devices 110, such as a hard drive and/or other computer readable media

having data recorded thereon. In some embodiments, the computer system 100 further includes one or more data retrieving device 118 for reading the data stored on the internal data storage devices 110.

5 The data retrieving device 118 may represent, for example, a floppy disk drive, a compact disk drive, a magnetic tape drive, etc. In some embodiments, the internal data storage device 110 is a removable computer readable medium such as a floppy disk, a compact disk, a magnetic tape, etc. containing control logic and/or data recorded thereon. The computer system 100 may advantageously include or be programmed by appropriate software for reading the control logic and/or the data from the data storage component once inserted in the data retrieving device.

10 The computer system 100 includes a display 120 which is used to display output to a computer user. It should also be noted that the computer system 100 can be linked to other computer systems 125a-c in a network or wide area network to provide centralized access to the computer system 100.

Software for accessing and processing the nucleotide sequences of the nucleic acid codes of the invention or the amino acid sequences of the polypeptide codes of the invention (such as search tools, compare tools, and modeling tools etc.) may reside in main memory 115 during execution.

In some embodiments, the computer system 100 may further comprise a sequence comparer for comparing the above-described nucleic acid codes of the invention or the polypeptide codes of the invention stored on a computer readable medium to reference nucleotide or polypeptide sequences stored on a computer readable medium. A "sequence comparer" refers to one or more programs which are implemented on the computer system 100 to compare a nucleotide or polypeptide sequence with other nucleotide or polypeptide sequences and/or compounds including but not limited to peptides, peptidomimetics, and chemicals stored within the data storage means. For example, the sequence comparer may compare the nucleotide sequences of nucleic acid codes of the invention or the amino acid sequences of the polypeptide codes of the invention stored on a computer readable medium to reference sequences stored on a computer readable medium to identify homologies, motifs implicated in biological function, or structural motifs. The various sequence comparer programs identified elsewhere in this patent specification are particularly contemplated for use in this aspect of the invention.

30 Figure 3 is a flow diagram illustrating one embodiment of a process 200 for comparing a new nucleotide or protein sequence with a database of sequences in order to determine the homology levels between the new sequence and the sequences in the database. The database of sequences can be a private database stored within the computer system 100, or a public database such as GENBANK, PIR OR SWISSPROT that is available through the Internet.

35 The process 200 begins at a start state 201 and then moves to a state 202 wherein the new sequence to be compared is stored to a memory in a computer system 100. As discussed above, the memory could be any type of memory, including RAM or an internal storage device.

The process 200 then moves to a state 204 wherein a database of sequences is opened for analysis and comparison. The process 200 then moves to a state 206 wherein the first sequence stored in the database is read into a memory on the computer. A comparison is then performed at a state 210 to determine if the first sequence is the same as the second sequence. It is important to note that this step is not limited to performing an exact comparison between the new sequence and the first sequence in the database. Well-known methods are known to those of skill in the art for comparing two nucleotide or protein sequences, even if they are not identical. For example, gaps can be introduced into one sequence in order to raise the homology level between the two tested sequences. The parameters that control whether gaps or other features are introduced into a sequence during comparison are normally entered by the user of the computer system.

Once a comparison of the two sequences has been performed at the state 210, a determination is made at a decision state 210 whether the two sequences are the same. Of course, the term "same" is not limited to sequences that are absolutely identical. Sequences that are within the homology parameters entered by the user will be marked as "same" in the process 200.

If a determination is made that the two sequences are the same, the process 200 moves to a state 214 wherein the name of the sequence from the database is displayed to the user. This state notifies the user that the sequence with the displayed name fulfills the homology constraints that were entered. Once the name of the stored sequence is displayed to the user, the process 200 moves to a decision state 218 wherein a determination is made whether more sequences exist in the database. If no more sequences exist in the database, then the process 200 terminates at an end state 220. However, if more sequences do exist in the database, then the process 200 moves to a state 224 wherein a pointer is moved to the next sequence in the database so that it can be compared to the new sequence. In this manner, the new sequence is aligned and compared with every sequence in the database.

It should be noted that if a determination had been made at the decision state 212 that the sequences were not homologous, then the process 200 would move immediately to the decision state 218 in order to determine if any other sequences were available in the database for comparison.

Accordingly, one aspect of the present invention is a computer system comprising a processor, a data storage device having stored thereon a nucleic acid code of the invention or a polypeptide code of the invention,. In some embodiments the computer system further comprises a data storage device having retrievably stored thereon reference nucleotide sequences or polypeptide sequences to be compared to the nucleic acid code of the invention or polypeptide code of the invention and a sequence comparer for conducting the comparison. For example, the sequence comparer may comprise a computer program which indicates polymorphisms. In other aspects of the computer system, the system further comprises an identifier which identifies features in said sequence. The sequence comparer may indicate a homology level between the sequences compared

or identify motifs implicated in biological function and structural motifs in the nucleic acid code of the invention and polypeptide codes of the invention or it may identify structural motifs in sequences which are compared to these nucleic acid codes and polypeptide codes. In some embodiments, the data storage device may have stored thereon the sequences of at least 2, 5, 10, 15, 20, 25, 30, 50, 75, 100, 150 or 200 of the nucleic acid codes of the invention or polypeptide codes of the invention.

Another aspect of the present invention is a method for determining the level of homology between a nucleic acid code of the invention and a reference nucleotide sequence, comprising the steps of reading the nucleic acid code and the reference nucleotide sequence through the use of a computer program which determines homology levels and determining homology between the nucleic acid code and the reference nucleotide sequence with the computer program. The computer program may be any of a number of computer programs for determining homology levels, including those specifically enumerated herein, including BLAST2N with the default parameters or with any modified parameters. The method may be implemented using the computer systems described above. The method may also be performed by reading 2, 5, 10, 15, 20, 25, 30, 50, 75, 100, 150 or 200 of the above described nucleic acid codes of the invention through the use of the computer program and determining homology between the nucleic acid codes and reference nucleotide sequences.

Figure 4 is a flow diagram illustrating one embodiment of a process 250 in a computer for determining whether two sequences are homologous. The process 250 begins at a start state 252 and then moves to a state 254 wherein a first sequence to be compared is stored to a memory. The second sequence to be compared is then stored to a memory at a state 256. The process 250 then moves to a state 260 wherein the first character in the first sequence is read and then to a state 262 wherein the first character of the second sequence is read. It should be understood that if the sequence is a nucleotide sequence, then the character would normally be either A, T, C, G or U. If the sequence is a protein sequence, then it should be in the single letter amino acid code so that the first and sequence sequences can be easily compared.

A determination is then made at a decision state 264 whether the two characters are the same. If they are the same, then the process 250 moves to a state 268 wherein the next characters in the first and second sequences are read. A determination is then made whether the next characters are the same. If they are, then the process 250 continues this loop until two characters are not the same. If a determination is made that the next two characters are not the same, the process 250 moves to a decision state 274 to determine whether there are any more characters either sequence to read.

If there aren't any more characters to read, then the process 250 moves to a state 276 wherein the level of homology between the first and second sequences is displayed to the user. The level of homology is determined by calculating the proportion of characters between the sequences

that were the same out of the total number of sequences in the first sequence. Thus, if every character in a first 100 nucleotide sequence aligned with a every character in a second sequence, the homology level would be 100%.

Alternatively, the computer program may be a computer program which compares the
5 nucleotide sequences of the nucleic acid codes of the present invention, to reference nucleotide sequences in order to determine whether the nucleic acid code of the invention differs from a reference nucleic acid sequence at one or more positions. Optionally such a program records the length and identity of inserted, deleted or substituted nucleotides with respect to the sequence of either the reference polynucleotide or the nucleic acid code of the invention. In one embodiment,
10 the computer program may be a program which determines whether the nucleotide sequences of the nucleic acid codes of the invention contain one or more single nucleotide polymorphisms (SNP) with respect to a reference nucleotide sequence. These single nucleotide polymorphisms may each comprise a single base substitution, insertion, or deletion.

Another embodiment of the present invention is a method for comparing a first sequence to
15 a reference sequence wherein the first sequence is selected from the group consisting of a cDNA code of SEQ ID NOs. 1-297 and a polypeptide code of SEQ ID NOs. 298-594 comprising the steps of reading the first sequence and the reference sequence through use of a computer program which compares sequences and determining differences between the first sequence and the reference sequence with the computer program. In some aspects of this embodiment, said step of determining
20 differences between the first sequence and the reference sequence comprises identifying polymorphisms.

Another aspect of the present invention is a method for determining the level of homology between a polypeptide code of the invention and a reference polypeptide sequence, comprising the steps of reading the polypeptide code of the invention and the reference polypeptide sequence
25 through use of a computer program which determines homology levels and determining homology between the polypeptide code and the reference polypeptide sequence using the computer program.

Accordingly, another aspect of the present invention is a method for determining whether a nucleic acid code of the invention differs at one or more nucleotides from a reference nucleotide sequence comprising the steps of reading the nucleic acid code and the reference nucleotide
30 sequence through use of a computer program which identifies differences between nucleic acid sequences and identifying differences between the nucleic acid code and the reference nucleotide sequence with the computer program. In some embodiments, the computer program is a program which identifies single nucleotide polymorphisms The method may be implemented by the computer systems described above and the method illustrated in Figure 4. The method may also be
35 performed by reading at least 2, 5, 10, 15, 20, 25, 30, 50, 75, 100, 150 or 200 of the nucleic acid codes of the invention and the reference nucleotide sequences through the use of the computer

program and identifying differences between the nucleic acid codes and the reference nucleotide sequences with the computer program.

Thus, another embodiment of the present invention is a method for comparing a first sequence to a reference sequence wherein the first sequence is selected from the group consisting of the nucleic acid codes of the present invention or the polypeptide codes of the present invention comprising the steps of reading the first sequence and the reference sequence through use of a computer program which compares sequences and determining differences between the first sequence and the reference sequence with the computer program. In some aspects of this embodiment, said step of determining differences between the first sequence and the reference sequence comprises identifying polymorphisms.

Another aspect of the present invention is a method for determining the level of identity between a first sequence and a reference sequence, wherein the first sequence is selected from the group consisting of the nucleic acid codes of the present invention or the polypeptide codes of the present invention, comprising the steps of reading the first sequence and the reference sequence through the use of a computer program which determines identity levels and determining identity between the first sequence and the reference sequence with the computer program.

In other embodiments the computer based system may further comprise an identifier for identifying features within the nucleotide sequences of the nucleic acid codes of the invention or the amino acid sequences of the polypeptide codes of the invention. An "identifier" refers to one or more programs which identifies certain features within the above-described nucleotide sequences of the nucleic acid codes of the invention or the amino acid sequences of the polypeptide codes of the invention. In one embodiment, the identifier may comprise a program which identifies an open reading frame in the cDNAs codes of the invention.

Another embodiment of the present invention is a method for identifying a feature in a sequence selected from the group consisting of the nucleic acid codes of the invention or the amino acid sequences of the polypeptide codes of the invention comprising the steps of reading the sequence through the use of a computer program which identifies features in sequences and identifying features in the sequence with said computer program. In one aspect of this embodiment, the computer program comprises a computer program which identifies open reading frames. In a further embodiment, the computer program comprises a program that identifies linear or structural motifs in a polypeptide sequence.

Figure 5 is a flow diagram illustrating one embodiment of an identifier process 300 for detecting the presence of a feature in a sequence. The process 300 begins at a start state 302 and then moves to a state 304 wherein a first sequence that is to be checked for features is stored to a memory 115 in the computer system 100. The process 300 then moves to a state 306 wherein a database of sequence features is opened. Such a database would include a list of each feature's attributes along with the name of the feature. For example, a feature name could be "Initiation

Codon" and the attribute would be "ATG". Another example would be the feature name "TAATAA Box" and the feature attribute would be "TAATAA". An example of such a database is produced by the University of Wisconsin Genetics Computer Group (www.gcg.com).

Once the database of features is opened at the state 306, the process 300 moves to a state 308 wherein the first feature is read from the database. A comparison of the attribute of the first feature with the first sequence is then made at a state 310. A determination is then made at a decision state 316 whether the attribute of the feature was found in the first sequence. If the attribute was found, then the process 300 moves to a state 318 wherein the name of the found feature is displayed to the user.

The process 300 then moves to a decision state 320 wherein a determination is made whether more features exist in the database. If no more features do exist, then the process 300 terminates at an end state 324. However, if more features do exist in the database, then the process 300 reads the next sequence feature at a state 326 and loops back to the state 310 wherein the attribute of the next feature is compared against the first sequence.

It should be noted, that if the feature attribute is not found in the first sequence at the decision state 316, the process 300 moves directly to the decision state 320 in order to determine if any more features exist in the database.

In another embodiment, the identifier may comprise a molecular modeling program which determines the 3-dimensional structure of the polypeptides codes of the invention. Such programs may use any methods known to those skilled in the art including methods based on homology-modeling, fold recognition and *ab initio* methods as described in Sternberg *et al.*, 1999, which disclosure is hereby incorporated by reference in its entirety. In some embodiments, the molecular modeling program identifies target sequences that are most compatible with profiles representing the structural environments of the residues in known three-dimensional protein structures. (See, e.g., Eisenberg *et al.*, U.S. Patent No. 5,436,850 issued July 25, 1995, which disclosure is hereby incorporated by reference in its entirety). In another technique, the known three-dimensional structures of proteins in a given family are superimposed to define the structurally conserved regions in that family. This protein modeling technique also uses the known three-dimensional structure of a homologous protein to approximate the structure of the polypeptide codes of the invention. (See e.g., Srinivasan, *et al.*, U.S. Patent No. 5,557,535 issued September 17, 1996, which disclosure is hereby incorporated by reference in its entirety). Conventional homology modeling techniques have been used routinely to build models of proteases and antibodies. (Sowdhamini *et al.*, (1997)). Comparative approaches can also be used to develop three-dimensional protein models when the protein of interest has poor sequence identity to template proteins. In some cases, proteins fold into similar three-dimensional structures despite having very weak sequence identities. For example, the three-dimensional structures of a number of helical cytokines fold in similar three-dimensional topology in spite of weak sequence homology.

The recent development of threading methods now enables the identification of likely folding patterns in a number of situations where the structural relatedness between target and template(s) is not detectable at the sequence level. Hybrid methods, in which fold recognition is performed using Multiple Sequence Threading (MST), structural equivalencies are deduced from the threading output using a distance geometry program DRAGON to construct a low resolution model, and a full-atom representation is constructed using a molecular modeling package such as QUANTA.

According to this 3-step approach, candidate templates are first identified by using the novel fold recognition algorithm MST, which is capable of performing simultaneous threading of multiple aligned sequences onto one or more 3-D structures. In a second step, the structural equivalencies obtained from the MST output are converted into interresidue distance restraints and fed into the distance geometry program DRAGON, together with auxiliary information obtained from secondary structure predictions. The program combines the restraints in an unbiased manner and rapidly generates a large number of low resolution model confirmations. In a third step, these low resolution model confirmations are converted into full-atom models and subjected to energy minimization using the molecular modeling package QUANTA. (See e.g., Aszódi *et al.*, (1997)).

The results of the molecular modeling analysis may then be used in rational drug design techniques to identify agents which modulate the activity of the polypeptide codes of the invention.

Accordingly, another aspect of the present invention is a method of identifying a feature within the nucleic acid codes of the invention or the polypeptide codes of the invention comprising reading the nucleic acid code(s) or the polypeptide code(s) through the use of a computer program which identifies features therein and identifying features within the nucleic acid code(s) or polypeptide code(s) with the computer program. In one embodiment, computer program comprises a computer program which identifies open reading frames. In a further embodiment, the computer program identifies linear or structural motifs in a polypeptide sequence. In another embodiment, the computer program comprises a molecular modeling program. The method may be performed by reading a single sequence or at least 2, 5, 10, 15, 20, 25, 30, 50, 75, 100, 150 or 200 of the nucleic acid codes of the invention or the polypeptide codes of the invention through the use of the computer program and identifying features within the nucleic acid codes or polypeptide codes with the computer program.

The nucleic acid codes of the invention or the polypeptide codes of the invention may be stored and manipulated in a variety of data processor programs in a variety of formats. For example, they may be stored as text in a word processing file, such as MicrosoftWORD or WORDPERFECT or as an ASCII file in a variety of database programs familiar to those of skill in the art, such as DB2, SYBASE, or ORACLE. In addition, many computer programs and databases may be used as sequence comparers, identifiers, or sources of reference nucleotide or polypeptide sequences to be compared to the nucleic acid codes of the invention or the polypeptide codes of the

invention. The following list is intended not to limit the invention but to provide guidance to programs and databases which are useful with the nucleic acid codes of the invention or the polypeptide codes of the invention. The programs and databases which may be used include, but are not limited to: MacPattern (EMBL), DiscoveryBase (Molecular Applications Group), GeneMine
5 (Molecular Applications Group), Look (Molecular Applications Group), MacLook (Molecular Applications Group), BLAST and BLAST2 (NCBI), BLASTN and BLASTX (Altschul *et al.*, 1990), FASTA (Pearson and Lipman, 1988), FASTDB (Brutlag *et al.*, 1990), Catalyst (Molecular Simulations Inc.), Catalyst/SHAPE (Molecular Simulations Inc.), Cerius2.DBAccess (Molecular Simulations Inc.), HypoGen (Molecular Simulations Inc.), Insight II, (Molecular Simulations Inc.),
10 Discover (Molecular Simulations Inc.), CHARMM (Molecular Simulations Inc.), Felix (Molecular Simulations Inc.), DelPhi, (Molecular Simulations Inc.), QuanteMM, (Molecular Simulations Inc.), Homology (Molecular Simulations Inc.), Modeler (Molecular Simulations Inc.), ISIS (Molecular Simulations Inc.), Quanta/Protein Design (Molecular Simulations Inc.), WebLab (Molecular Simulations Inc.), WebLab Diversity Explorer (Molecular Simulations Inc.), Gene Explorer
15 (Molecular Simulations Inc.), SeqFold (Molecular Simulations Inc.), the EMBL/Swissprotein database, the MDL Available Chemicals Directory database, the MDL Drug Data Report data base, the Comprehensive Medicinal Chemistry database, Derwent's World Drug Index database, the BioByteMasterFile database, the Genbank database, and the Genseqn database. Many other programs and data bases would be apparent to one of skill in the art given the present disclosure:

20 Motifs which may be detected using the above programs include sequences encoding leucine zippers, helix-turn-helix motifs, glycosylation sites, ubiquitination sites, alpha helices, and beta sheets, signal sequences encoding signal peptides which direct the secretion of the encoded proteins, sequences implicated in transcription regulation such as homeoboxes, acidic stretches, enzymatic active sites, substrate binding sites, and enzymatic cleavage sites.

25 CONCLUSION

As discussed above, the GENSET polynucleotides and polypeptides of the present invention or fragments thereof can be used for various purposes. The polynucleotides can be used to express recombinant protein for analysis, characterization or therapeutic use; as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a
30 particular stage of tissue differentiation or development or in disease states); as molecular weight markers on Southern gels; as chromosome markers or tags (when labeled) to identify chromosomes or to map related gene positions; as a reagent (including a labeled reagent) in assays designed to quantitatively determine levels of GENSET expression in biological samples; to compare with endogenous DNA sequences in patients to identify potential genetic disorders; as probes to
35 hybridize and thus discover novel, related DNA sequences; as a source of information to derive PCR primers for genetic fingerprinting; for selecting and making oligomers for attachment to a

"gene chip" or other support, including for examination for expression patterns; to raise anti-protein antibodies using DNA immunization techniques; and as an antigen to raise anti-DNA antibodies or elicit another immune response. Where the polynucleotide encodes a protein which binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the polynucleotide can also be used in interaction trap assays (such as, for example, that described in Gyuris *et al.*, (1993) to identify polynucleotides encoding the other protein with which binding occurs or to identify inhibitors of the binding interaction.

The proteins or polypeptides provided by the present invention can similarly be used in assays to determine biological activity, including in a panel of multiple proteins for high-throughput screening; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its receptor) in biological fluids; as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state); and, of course, to isolate correlative receptors or ligands. Where the protein binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the protein can be used to identify the other protein with which binding occurs or to identify inhibitors of the binding interaction. Proteins involved in these binding interactions can also be used to screen for peptide or small molecule inhibitors or agonists of the binding interaction.

Any or all of these research utilities are capable of being developed into reagent grade or kit format for commercialization as research products.

Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include without limitation "Molecular Cloning; A Laboratory Manual", 2d ed., Cole Spring Harbor Laboratory Press, Sambrook, J., E.F. Fritsch and T. Maniatis eds., 1989, and "Methods in Enzymology; Guide to Molecular Cloning Techniques", Academic Press, Berger and Kimmel eds., 1987, which disclosures are hereby incorporated by reference in their entireties.

Polynucleotides and proteins of the present invention can also be used as nutritional sources or supplements. Such uses include without limitation use as a protein or amino acid supplement, use as a carbon source, use as a nitrogen source and use as a source of carbohydrate. In such cases the protein or polynucleotide of the invention can be added to the feed of a particular organism or can be administered as a separate solid or liquid preparation, such as in the form of powder, pills, solutions, suspensions or capsules. In the case of microorganisms, the protein or polynucleotide of the invention can be added to the medium in or on which the microorganism is cultured.

Although this invention has been described in terms of certain preferred embodiments, other embodiments which will be apparent to those of ordinary skill in the art in view of the disclosure herein are also within the scope of this invention. Accordingly, the scope of the invention is intended to be defined only by reference to the appended claims.

EXAMPLES

Preparation of Antibody Compositions to GENSET proteins

Substantially pure protein or polypeptide is isolated from transfected or transformed cells containing an expression vector encoding a GENSET protein or a portion thereof. The concentration of protein in the final preparation is adjusted, for example, by concentration on an Amicon filter device, to the level of a few micrograms/ml. Monoclonal or polyclonal antibody to the protein can then be prepared as follows:

A. Monoclonal Antibody Production by Hybridoma Fusion

Monoclonal antibody to epitopes in the GENSET protein or a portion thereof can be prepared from murine hybridomas according to the classical method of Kohler and Milstein, (1975) or derivative methods thereof. Also see Harlow and Lane. (1988)..

Briefly, a mouse is repetitively inoculated with a few micrograms of the GENSET protein or a portion thereof over a period of a few weeks. The mouse is then sacrificed, and the antibody producing cells of the spleen isolated. The spleen cells are fused by means of polyethylene glycol with mouse myeloma cells, and the excess unfused cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate where growth of the culture is continued. Antibody-producing clones are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures, such as ELISA, as originally described by Engvall, (1980), which disclosure is hereby incorporated by reference in its entirety, and derivative methods thereof. Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Davis, *et al.* (1986) Section 21-2.

B. Polyclonal Antibody Production by Immunization

Polyclonal antiserum containing antibodies to heterogeneous epitopes in the GENSET protein or a portion thereof can be prepared by immunizing suitable non-human animal with the GENSET protein or a portion thereof, which can be unmodified or modified to enhance immunogenicity. A suitable non-human animal is preferably a non-human mammal is selected, usually a mouse, rat, rabbit, goat, or horse. Alternatively, a crude preparation which has been enriched for GENSET concentration can be used to generate antibodies. Such proteins, fragments or preparations are introduced into the non-human mammal in the presence of an appropriate adjuvant (e.g. aluminum hydroxide, RIBI, etc.) which is known in the art. In addition the protein, fragment or preparation can be pretreated with an agent which will increase antigenicity, such agents are known in the art and include, for example, methylated bovine serum albumin (mBSA), bovine serum albumin (BSA), Hepatitis B surface antigen, and keyhole limpet hemocyanin (KLH).

Serum from the immunized animal is collected, treated and tested according to known procedures. If the serum contains polyclonal antibodies to undesired epitopes, the polyclonal antibodies can be purified by immunoaffinity chromatography.

Effective polyclonal antibody production is affected by many factors related both to the antigen and the host species. Also, host animals vary in response to site of inoculations and dose, with both inadequate or excessive doses of antigen resulting in low titer antisera. Small doses (ng level) of antigen administered at multiple intradermal sites appears to be most reliable. Techniques for producing and processing polyclonal antisera are known in the art. An effective immunization protocol for rabbits can be found in Vaitukaitis *et al.* (1971), which disclosure is hereby incorporated by reference in its entirety.

Booster injections can be given at regular intervals, and antiserum harvested when antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. See, for example, Ouchterlony *et al.*, (1973), which disclosure is hereby incorporated by reference in its entirety. Plateau concentration of antibody is usually in the range of 0.1 to 0.2 mg/ml of serum (about 12 uM). Affinity of the antisera for the antigen is determined by preparing competitive binding curves, as described, for example, by Fisher (1980), which disclosure is hereby incorporated by reference in its entirety.

Antibody preparations prepared according to either the monoclonal or the polyclonal protocol are useful in quantitative immunoassays which determine concentrations of antigen-bearing substances in biological samples; they are also used semi-quantitatively or qualitatively to identify the presence of antigen in a biological sample. The antibodies may also be used in therapeutic compositions for killing cells expressing the protein or reducing the levels of the protein in the body.

Biological assays

25 Assaying GENSET Secreted Proteins to Determine Whether they Bind to the Cell Surface

The secreted proteins encoded by the GENSET cDNAs, preferably the proteins of SEQ ID NOs: 242-272 and 274-384, or fragments thereof are cloned into expression vectors. The proteins are purified by size, charge, immunochromatography or other techniques familiar to those skilled in the art. Following purification, the proteins are labeled using techniques known to those skilled in the art. The labeled proteins are incubated with cells or cell lines derived from a variety of organs or tissues to allow the proteins to bind to any receptor present on the cell surface. Following the incubation, the cells are washed to remove non-specifically bound protein. The labeled proteins are detected by autoradiography. Alternatively, unlabeled proteins may be incubated with the cells and detected with antibodies having a detectable label, such as a fluorescent molecule, attached thereto.

35 Specificity of cell surface binding may be analyzed by conducting a competition analysis in which various amounts of unlabeled protein are incubated along with the labeled protein. The

amount of labeled protein bound to the cell surface decreases as the amount of competitive unlabeled protein increases. As a control, various amounts of an unlabeled protein unrelated to the labeled protein is included in some binding reactions. The amount of labeled protein bound to the cell surface does not decrease in binding reactions containing increasing amounts of unrelated
5 unlabeled protein, indicating that the protein encoded by the cDNA binds specifically to the cell surface.

As discussed herein, secreted proteins have been shown to have a number of important physiological effects and, consequently, represent a valuable therapeutic resource. The secreted proteins encoded by the cDNAs or fragments thereof made using any of the methods described
10 therein may be evaluated to determine their physiological activities as described below.

Assaying GENSET proteins or Fragments Thereof for Cytokine, Cell Proliferation or Cell Differentiation Activity

Secreted proteins may act as cytokines or may affect cellular proliferation or differentiation. Many protein factors discovered to date, including all known cytokines, have exhibited activity in
15 one or more factor dependent cell proliferation assays, and hence the assays serve as a convenient confirmation of cytokine activity. The activity of a protein of the present invention is evidenced by any one of a number of routine factor dependent cell proliferation assays for cell lines including, without limitation, 32D, DA2, DA1G, T10, B9, B9/11, BaF3, MC9/G, M+ (preB M+), 2E8, RB5, DA1, 123, T1165, HT2, CTLL2, TF-1, Mo7c and CMK. The proteins encoded by the cDNAs of
20 the invention or fragments thereof may be evaluated for their ability to regulate T cell or thymocyte proliferation in assays such as those described above or in the following references, which are incorporated herein by reference: Current Protocols in Immunology, Ed. by J.E. Coligan *et al.*, Greene Publishing Associates and Wiley-Interscience; Takai *et al. J. Immunol.* 137:3494-3500, 1986. Bertagnolli *et al. J. Immunol.* 145:1706-1712, 1990. Bertagnolli *et al., Cellular Immunology*
25 133:327-341, 1991. Bertagnolli, *et al. J. Immunol.* 149:3778-3783, 1992; Bowman *et al., J. Immunol.* 152:1756-1761, 1994.

In addition, numerous assays for cytokine production and/or the proliferation of spleen cells, lymph node cells and thymocytes are known. These include the techniques disclosed in Current Protocols in Immunology. J.E. Coligan *et al.* Eds., Vol 1 pp. 3.12.1-3.12.14 John Wiley and
30 Sons, Toronto. 1994; and Schreiber, R.D. Current Protocols in Immunology, *supra* Vol 1 pp. 6.8.1-6.8.8, John Wiley and Sons, Toronto. 1994.

The proteins encoded by the cDNAs of the invention may also be assayed for the ability to regulate the proliferation and differentiation of hematopoietic or lymphopoietic cells. Many assays for such activity are familiar to those skilled in the art, including the assays in the following
35 references, which are incorporated herein by reference: Bottomly, K., Davis, L.S. and Lipsky, P.E., Measurement of Human and Murine Interleukin 2 and Interleukin 4, Current Protocols in

Immunology, J.E. Coligan *et al.* Eds. Vol 1 pp. 6.3.1-6.3.12, John Wiley and Sons, Toronto. 1991; deVries *et al.*, *J. Exp. Med.* 173:1205-1211, 1991; Moreau *et al.*, *Nature* 36:690-692, 1988; Greenberger *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 80:2931-2938, 1983; Nordan, R., Measurement of Mouse and Human Interleukin 6 Current Protocols in Immunology. J.E. Coligan *et al.* Eds. Vol 1
 5 pp. 6.6.1-6.6.5, John Wiley and Sons, Toronto. 1991; Smith *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 83:1857-1861, 1986; Bennett, F., Giannotti, J., Clark, S.C. and Turner, K.J., Measurement of Human Interleukin 11 Current Protocols in Immunology. J.E. Coligan *et al.* Eds. Vol 1 pp. 6.15.1 John Wiley and Sons, Toronto. 1991; Ciarletta, A., Giannotti, J., Clark, S.C. and Turner, K.J., Measurement of Mouse and Human Interleukin 9 Current Protocols in Immunology. J.E. Coligan *et al.*,
 10 *al.*, Eds. Vol 1 pp. 6.13.1, John Wiley and Sons, Toronto. 1991.

The proteins encoded by the cDNAs of the invention may also be assayed for their ability to regulate T-cell responses to antigens. Many assays for such activity are familiar to those skilled in the art, including the assays described in the following references, which are incorporated herein by reference: Chapter 3 (*In vitro* Assays for Mouse Lymphocyte Function), Chapter 6 (Cytokines and
 15 Their Cellular Receptors) and Chapter 7, (Immunologic Studies in Humans) in Current Protocols in Immunology, J.E. Coligan *et al.* Eds. Greene Publishing Associates and Wiley-Interscience; Weinberger *et al.*, *Proc. Natl. Acad. Sci. USA* 77:6091-6095, 1980; Weinberger *et al.*, *Eur. J. Immunol.* 11:405-411, 1981; Takai *et al.*, *J. Immunol.* 137:3494-3500, 1986; Takai *et al.*, *J. Immunol.* 140:508-512, 1988.

20 Those proteins which exhibit cytokine, cell proliferation, or cell differentiation activity may then be formulated as pharmaceuticals and used to treat clinical conditions in which induction of cell proliferation or differentiation is beneficial. Alternatively, as described in more detail below, genes encoding these proteins or nucleic acids regulating the expression of these proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as
 25 desired.

Assaying GENSET proteins or Fragments Thereof for Activity as Immune System Regulators

The proteins encoded by the cDNAs of the invention may also be evaluated for their effects as immune regulators. For example, the proteins may be evaluated for their activity to influence thymocyte or splenocyte cytotoxicity. Numerous assays for such activity are familiar to those
 30 skilled in the art including the assays described in the following references, which are incorporated herein by reference: Chapter 3 (*In vitro* Assays for Mouse Lymphocyte Function 3.1-3.19) and Chapter 7 (Immunologic studies in Humans) in Current Protocols in Immunology, J.E. Coligan *et al.* Eds, Greene Publishing Associates and Wiley-Interscience; Herrmann *et al.*, *Proc. Natl. Acad. Sci. USA* 78:2488-2492, 1981; Herrmann *et al.*, *J. Immunol.* 128:1968-1974, 1982; Handa *et al.*, *J. Immunol.* 135:1564-1572, 1985; Takai *et al.*, *J. Immunol.* 137:3494-3500, 1986; Takai *et al.*, *J. Immunol.* 140:508-512, 1988; Herrmann *et al.*, *Proc. Natl. Acad. Sci. USA* 78:2488-2492, 1981;
 35

Herrmann *et al.*, *J. Immunol.* 128:1968-1974, 1982; Handa *et al.*, *J. Immunol.* 135:1564-1572, 1985; Takai *et al.*, *J. Immunol.* 137:3494-3500, 1986; Bowman *et al.*, *J. Virology* 61:1992-1998; Takai *et al.*, *J. Immunol.* 140:508-512, 1988; Bertagnolli *et al.*, *Cellular Immunology* 133:327-341, 1991; Brown *et al.*, *J. Immunol.* 153:3079-3092, 1994.

5 The proteins encoded by the cDNAs of the invention may also be evaluated for their effects on T-cell dependent immunoglobulin responses and isotype switching. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Maliszewski, *J. Immunol.* 144:3028-3033, 1990; Mond, J.J. and Brunswick, M Assays for B Cell Function: *In vitro* Antibody Production, Vol
10 1 pp. 3.8.1-3.8.16 in Current Protocols in Immunology. J.E. Coligan *et al* Eds., John Wiley and Sons, Toronto. 1994.

 The proteins encoded by the cDNAs of the invention may also be evaluated for their effect on immune effector cells, including their effect on Th1 cells and cytotoxic lymphocytes. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the
15 following references, which are incorporated herein by reference: Chapter 3 (*In vitro* Assays for Mouse Lymphocyte Function 3.1-3.19) and Chapter 7 (Immunologic Studies in Humans) in Current Protocols in Immunology, J.E. Coligan *et al.* Eds., Greene Publishing Associates and Wiley-Interscience; Takai *et al.*, *J. Immunol.* 137:3494-3500, 1986; Takai *et al.*; *J. Immunol.* 140:508-512, 1988; Bertagnolli *et al.*, *J. Immunol.* 149:3778-3783, 1992.

20 The proteins encoded by the cDNAs of the invention may also be evaluated for their effect on dendritic cell mediated activation of naive T-cells. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Guery *et al.*, *J. Immunol.* 134:536-544, 1995; Inaba *et al.*, *Journal of Experimental Medicine* 173:549-559, 1991; Macatonia *et al.*, *Journal of Immunology*
25 154:5071-5079, 1995; Porgador *et al.*, *Journal of Experimental Medicine* 182:255-260, 1995; Nair *et al.*, *Journal of Virology* 67:4062-4069, 1993; Huang *et al.*, *Science* 264:961-965, 1994; Macatonia *et al.*, *Journal of Experimental Medicine* 169:1255-1264, 1989; Bhardwaj *et al.*, *Journal of Clinical Investigation* 94:797-807, 1994; and Inaba *et al.*, *Journal of Experimental Medicine* 172:631-640, 1990.

30 The proteins encoded by the cDNAs of the invention may also be evaluated for their influence on the lifetime of lymphocytes. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Darzynkiewicz *et al.*, *Cytometry* 13:795-808, 1992; Gorczyca *et al.*, *Leukemia*
35 7:659-670, 1993; Gorczyca *et al.*, *Cancer Research* 53:1945-1951, 1993; Itoh *et al.*, *Cell* 66:233-243, 1991; Zacharchuk, *Journal of Immunology* 145:4037-4045, 1990; Zamai *et al.*, *Cytometry* 14:891-897, 1993; Gorczyca *et al.*, *International Journal of Oncology* 1:639-648, 1992.

Assays for proteins that influence early steps of T-cell commitment and development include, without limitation, those described in: Antica *et al.*, Blood 84:111-117, 1994; Fine *et al.*, Cellular immunology 155:111-122, 1994; Galy *et al.*, Blood 85:2770-2778, 1995; Toki *et al.*, Proc. Nat. Acad Sci. USA 88:7548-7551, 1991.

5 Those proteins which exhibit activity as immune system regulators activity may then be formulated as pharmaceuticals and used to treat clinical conditions in which regulation of immune activity is beneficial. For example, the protein may be useful in the treatment of various immune deficiencies and disorders (including severe combined immunodeficiency (SCID)), e.g., in regulating (up or down) growth and proliferation of T and/or B lymphocytes, as well as effecting
10 the cytolytic activity of NK cells and other cell populations. These immune deficiencies may be genetic or be caused by viral (e.g., HIV) as well as bacterial or fungal infections, or may result from autoimmune disorders. More specifically, infectious diseases caused by viral, bacterial, fungal or other infection may be treatable using a protein of the present invention, including infections by HIV, hepatitis viruses, herpesviruses, mycobacteria, Leishmania spp., malaria spp. and various
15 fungal infections such as candidiasis. Of course, in this regard, a protein of the present invention may also be useful where a boost to the immune system generally may be desirable, i.e., in the treatment of cancer.

Autoimmune disorders which may be treated using a protein of the present invention include, for example, connective tissue disease, multiple sclerosis, systemic lupus erythematosus,
20 rheumatoid arthritis, autoimmune pulmonary inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent diabetes mellitis, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. Such a protein of the present invention may also to be useful in the treatment of allergic reactions and conditions, such as asthma (particularly allergic asthma) or other respiratory problems. Other conditions, in which immune suppression is desired
25 (including, for example, organ transplantation), may also be treatable using a protein of the present invention.

Using the proteins of the invention it may also be possible to regulate immune responses, in a number of ways. Down regulation may be in the form of inhibiting or blocking an immune response already in progress or may involve preventing the induction of an immune response. The
30 functions of activated T-cells may be inhibited by suppressing T cell responses or by inducing specific tolerance in T cells, or both. Immunosuppression of T cell responses is generally an active, non-antigen-specific, process which requires continuous exposure of the T cells to the suppressive agent. Tolerance, which involves inducing non-responsiveness or anergy in T cells, is distinguishable from immunosuppression in that it is generally antigen-specific and persists after
35 exposure to the tolerizing agent has ceased. Operationally, tolerance can be demonstrated by the lack of a T cell response upon reexposure to specific antigen in the absence of the tolerizing agent.

Down regulating or preventing one or more antigen functions (including without limitation B lymphocyte antigen functions (such as, for example, B7)), e.g., preventing high level lymphokine synthesis by activated T cells, will be useful in situations of tissue, skin and organ transplantation and in graft-versus-host disease (GVHD). For example, blockage of T cell function should result in reduced tissue destruction in tissue transplantation. Typically, in tissue transplants, rejection of the transplant is initiated through its recognition as foreign by T cells, followed by an immune reaction that destroys the transplant. The administration of a molecule which inhibits or blocks interaction of a B7 lymphocyte antigen with its natural ligand(s) on immune cells (such as a soluble, monomeric form of a peptide having B7-2 activity alone or in conjunction with a monomeric form of a peptide having an activity of another B lymphocyte antigen (e.g., B7-1, B7-3) or blocking antibody), prior to transplantation can lead to the binding of the molecule to the natural ligand(s) on the immune cells without transmitting the corresponding costimulatory signal. Blocking B lymphocyte antigen function in this matter prevents cytokine synthesis by immune cells, such as T cells, and thus acts as an immunosuppressant. Moreover, the lack of costimulation may also be sufficient to anergize the T cells, thereby inducing tolerance in a subject. Induction of long-term tolerance by B lymphocyte antigen-blocking reagents may avoid the necessity of repeated administration of these blocking reagents. To achieve sufficient immunosuppression or tolerance in a subject, it may also be necessary to block the function of a combination of B lymphocyte antigens.

The efficacy of particular blocking reagents in preventing organ transplant rejection or GVHD can be assessed using animal models that are predictive of efficacy in humans. Examples of appropriate systems which can be used include allogeneic cardiac grafts in rats and xenogeneic pancreatic islet cell grafts in mice, both of which have been used to examine the immunosuppressive effects of CTLA4Ig fusion proteins *in vivo* as described in Lenschow *et al.*, Science 257:789-792 (1992) and Turka *et al.*, Proc. Natl. Acad. Sci USA, 89:11102-11105 (1992). In addition, murine models of GVHD (see Paul ed., Fundamental Immunology, Raven Press, New York, 1989, pp. 846-847) can be used to determine the effect of blocking B lymphocyte antigen function *in vivo* on the development of that disease.

Blocking antigen function may also be therapeutically useful for treating autoimmune diseases. Many autoimmune disorders are the result of inappropriate activation of T cells that are reactive against self tissue and which promote the production of cytokines and autoantibodies involved in the pathology of the diseases. Preventing the activation of autoreactive T cells may reduce or eliminate disease symptoms. Administration of reagents which block costimulation of T cells by disrupting receptor ligand interactions of B lymphocyte antigens can be used to inhibit T cell activation and prevent production of autoantibodies or T cell-derived cytokines which may be involved in the disease process. Additionally, blocking reagents may induce antigen-specific tolerance of autoreactive T cells which could lead to long-term relief from the disease. The efficacy of blocking reagents in preventing or alleviating autoimmune disorders can be determined using a

number of well-characterized animal models of human autoimmune diseases. Examples include murine experimental autoimmune encephalitis, systemic lupus erythematosis in MRL/pr/pr mice or NZB hybrid mice, murine autoimmune collagen arthritis, diabetes mellitus in OD mice and BB rats, and murine experimental myasthenia gravis (see Paul ed., Fundamental Immunology, Raven Press, New York, 1989, pp. 840-856).

Upregulation of an antigen function (preferably a B lymphocyte antigen function), as a means of up regulating immune responses, may also be useful in therapy. Upregulation of immune responses may be in the form of enhancing an existing immune response or eliciting an initial immune response. For example, enhancing an immune response through stimulating B lymphocyte antigen function may be useful in cases of viral infection. In addition, systemic viral diseases such as influenza, the common cold, and encephalitis might be alleviated by the administration of stimulatory form of B lymphocyte antigens systemically.

Alternatively, anti-viral immune responses may be enhanced in an infected patient by removing T cells from the patient, costimulating the T cells *in vitro* with viral antigen-pulsed APCs either expressing a peptide of the present invention or together with a stimulatory form of a soluble peptide of the present invention and reintroducing the *in vitro* activated T cells into the patient. The infected cells would now be capable of delivering a costimulatory signal to T cells *in vivo*, thereby activating the T cells.

In another application, up regulation or enhancement of antigen function (preferably B lymphocyte antigen function) may be useful in the induction of tumor immunity. Tumor cells (e.g., sarcoma, melanoma, lymphoma, leukemia, neuroblastoma, carcinoma) transfected with a nucleic acid encoding at least one peptide of the present invention can be administered to a subject to overcome tumor-specific tolerance in the subject. If desired, the tumor cell can be transfected to express a combination of peptides. For example, tumor cells obtained from a patient can be transfected *ex vivo* with an expression vector directing the expression of a peptide having B7-2-like activity alone, or in conjunction with a peptide having B7-1-like activity and/or B7-3-like activity. The transfected tumor cells are returned to the patient to result in expression of the peptides on the surface of the transfected cell. Alternatively, gene therapy techniques can be used to target a tumor cell for transfection *in vivo*.

The presence of the peptide of the present invention having the activity of a B lymphocyte antigen(s) on the surface of the tumor cell provides the necessary costimulation signal to T cells to induce a T cell mediated immune response against the transfected tumor cells. In addition, tumor cells which lack MHC class I or MHC class II molecules, or which fail to reexpress sufficient amounts of MHC class I or MHC class II molecules, can be transfected with nucleic acids encoding all or a fragment of (e.g., a cytoplasmic-domain truncated fragment) of an MHC class I α chain protein and β_2 microglobulin protein or an MHC class II α chain protein and an MHC class II β chain protein to thereby express MHC class I or MHC class II proteins on the cell surface.

Expression of the appropriate class II or class II MHC in conjunction with a peptide having the activity of a B lymphocyte antigen (e.g., B7-1, B7-2, B7-3) induces a T cell mediated immune response against the transfected tumor cell. Optionally, a gene encoding an antisense construct which blocks expression of an MHC class II associated protein, such as the invariant chain, can also be cotransfected with a DNA encoding a peptide having the activity of a B lymphocyte antigen to promote presentation of tumor associated antigens and induce tumor specific immunity. Thus, the induction of a T cell mediated immune response in a human subject may be sufficient to overcome tumor-specific tolerance in the subject. Alternatively, as described in more detail below, genes encoding these proteins or nucleic acids regulating the expression of these proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

Assaying GENSET proteins or Fragments Thereof for Hematopoiesis Regulating Activity

The proteins encoded by the cDNAs of the invention or fragments thereof may also be evaluated for their hematopoiesis regulating activity. For example, the effect of the proteins on embryonic stem cell differentiation may be evaluated. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Johansson *et al.* *Cellular Biology* 15:141-151, 1995; Keller *et al.*, *Molecular and Cellular Biology* 13:473-486, 1993; McClanahan *et al.*, *Blood* 81:2903-2915, 1993.

The proteins encoded by the cDNAs of the invention or fragments thereof may also be evaluated for their influence on the lifetime of stem cells and stem cell differentiation. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Freshney, M.G. Methylcellulose Colony Forming Assays, in Culture of Hematopoietic Cells. R.I. Freshney, *et al.* Eds. pp. 265-268, Wiley-Liss, Inc., New York, NY. 1994; Hirayama *et al.*, *Proc. Natl. Acad. Sci. USA* 89:5907-5911, 1992; McNiece, I.K. and Briddell, R.A. Primitive Hematopoietic Colony Forming Cells with High Proliferative Potential, in Culture of Hematopoietic Cells. R.I. Freshney, *et al.* eds. Vol pp. 23-39, Wiley-Liss, Inc., New York, NY. 1994; Neben *et al.*, *Experimental Hematology* 22:353-359, 1994; Ploemacher, R.E. Cobblestone Area Forming Cell Assay, In Culture of Hematopoietic Cells. R.I. Freshney, *et al.* Eds. pp. 1-21, Wiley-Liss, Inc., New York, NY. 1994; Spooncer, E., Dexter, M. and Allen, T. Long Term Bone Marrow Cultures in the Presence of Stromal Cells, in Culture of Hematopoietic Cells. R.I. Freshney, *et al.* Eds. pp. 163-179, Wiley-Liss, Inc., New York, NY. 1994; and Sutherland, H.J. Long Term Culture Initiating Cell Assay, in Culture of Hematopoietic Cells. R.I. Freshney, *et al.* Eds. pp. 139-162, Wiley-Liss, Inc., New York, NY. 1994.

Those proteins which exhibit hematopoiesis regulatory activity may then be formulated as pharmaceuticals and used to treat clinical conditions in which regulation of hematopoiesis is

beneficial. For example, a protein of the present invention may be useful in regulation of hematopoiesis and, consequently, in the treatment of myeloid or lymphoid cell deficiencies. Even marginal biological activity in support of colony forming cells or of factor-dependent cell lines indicates involvement in regulating hematopoiesis, e.g. in supporting the growth and proliferation of erythroid progenitor cells alone or in combination with other cytokines, thereby indicating utility, for example, in treating various anemias or for use in conjunction with irradiation/chemotherapy to stimulate the production of erythroid precursors and/or erythroid cells; in supporting the growth and proliferation of myeloid cells such as granulocytes and monocytes/macrophages (i.e., traditional CSF activity) useful, for example, in conjunction with chemotherapy to prevent or treat consequent myelo-suppression; in supporting the growth and proliferation of megakaryocytes and consequently of platelets thereby allowing prevention or treatment of various platelet disorders such as thrombocytopenia, and generally for use in place of or complimentary to platelet transfusions; and/or in supporting the growth and proliferation of hematopoietic stem cells which are capable of maturing to any and all of the above-mentioned hematopoietic cells and therefore find therapeutic utility in various stem cell disorders (such as those usually treated with transplantation, including, without limitation, aplastic anemia and paroxysmal nocturnal hemoglobinuria), as well as in repopulating the stem cell compartment post irradiation/chemotherapy, either in-vivo or ex-vivo (i.e., in conjunction with bone marrow transplantation or with peripheral progenitor cell transplantation (homologous or heterologous)) as normal cells or genetically manipulated for gene therapy. Alternatively, as described in more detail below, genes encoding these proteins or nucleic acids regulating the expression of these proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

Assaying GENSET proteins or Fragments Thereof for Regulation of Tissue Growth

The proteins encoded by the cDNAs of the invention or fragments thereof may also be evaluated for their effect on tissue growth. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in International Patent Publication No. WO95/16035, International Patent Publication No. WO95/05846 and International Patent Publication No. WO91/07491, which are incorporated herein by reference.

Assays for wound healing activity include, without limitation, those described in: Winter, Epidermal Wound Healing, pps. 71-112 (Maibach, H1 and Rovee, DT, eds.), Year Book Medical Publishers, Inc., Chicago, as modified by Eaglstein and Mertz, J. Invest. Dermatol 71:382-84 (1978) which are incorporated herein by reference.

Those proteins which are involved in the regulation of tissue growth may then be formulated as pharmaceuticals and used to treat clinical conditions in which regulation of tissue growth is beneficial. For example, a protein of the present invention also may have utility in compositions used for bone, cartilage, tendon, ligament and/or nerve tissue growth or regeneration,

as well as for wound healing and tissue repair and replacement, and in the treatment of burns, incisions and ulcers.

A protein of the present invention, which induces cartilage and/or bone growth in circumstances where bone is not normally formed, has application in the healing of bone fractures and cartilage damage or defects in humans and other animals. Such a preparation employing a protein of the invention may have prophylactic use in closed as well as open fracture reduction and also in the improved fixation of artificial joints. De novo bone formation induced by an osteogenic agent contributes to the repair of congenital, trauma induced, or oncologic resection induced craniofacial defects, and also is useful in cosmetic plastic surgery.

10 A protein of this invention may also be used in the treatment of periodontal disease, and in other tooth repair processes. Such agents may provide an environment to attract bone-forming cells, stimulate growth of bone-forming cells or induce differentiation of progenitors of bone-forming cells. A protein of the invention may also be useful in the treatment of osteoporosis or osteoarthritis, such as through stimulation of bone and/or cartilage repair or by blocking
15 inflammation or processes of tissue destruction (collagenase activity, osteoclast activity, etc.) mediated by inflammatory processes.

Another category of tissue regeneration activity that may be attributable to the protein of the present invention is tendon/ligament formation. A protein of the present invention, which induces tendon/ligament-like tissue or other tissue formation in circumstances where such tissue is not
20 normally formed, has application in the healing of tendon or ligament tears, deformities and other tendon or ligament defects in humans and other animals. Such a preparation employing a tendon/ligament-like tissue inducing protein may have prophylactic use in preventing damage to tendon or ligament tissue, as well as use in the improved fixation of tendon or ligament to bone or other tissues, and in repairing defects to tendon or ligament tissue. De novo tendon/ligament-like
25 tissue formation induced by a composition of the present invention contributes to the repair of congenital, trauma induced, or other tendon or ligament defects of other origin, and is also useful in cosmetic plastic surgery for attachment or repair of tendons or ligaments. The compositions of the present invention may provide an environment to attract tendon- or ligament-forming cells, stimulate growth of tendon- or ligament-forming cells, induce differentiation of progenitors of
30 tendon- or ligament-forming cells, or induce growth of tendon/ligament cells or progenitors *ex vivo* for return *in vivo* to effect tissue repair. The compositions of the invention may also be useful in the treatment of tendinitis, carpal tunnel syndrome and other tendon or ligament defects. The compositions may also include an appropriate matrix and/or sequestering agent as a carrier as is well known in the art.

35 The protein of the present invention may also be useful for proliferation of neural cells and for regeneration of nerve and brain tissue, i.e., for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders, which involve

degeneration, death or trauma to neural cells or nerve tissue. More specifically, a protein may be used in the treatment of diseases of the peripheral nervous system, such as peripheral nerve injuries, peripheral neuropathy and localized neuropathies, and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-
5 Drager syndrome. Further conditions which may be treated in accordance with the present invention include mechanical and traumatic disorders, such as spinal cord disorders, head trauma and cerebrovascular diseases such as stroke. Peripheral neuropathies resulting from chemotherapy or other medical therapies may also be treatable using a protein of the invention.

Proteins of the invention may also be useful to promote better or faster closure of non-
10 healing wounds, including without limitation pressure ulcers, ulcers associated with vascular insufficiency, surgical and traumatic wounds, and the like.

It is expected that a protein of the present invention may also exhibit activity for generation or regeneration of other tissues, such as organs (including, for example, pancreas, liver, intestine, kidney, skin, endothelium) muscle (smooth, skeletal or cardiac) and vascular (including vascular
15 endothelium) tissue, or for promoting the growth of cells comprising such tissues. Part of the desired effects may be by inhibition or modulation of fibrotic scarring to allow normal tissue to generate. A protein of the invention may also exhibit angiogenic activity.

A protein of the present invention may also be useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, and conditions resulting
20 from systemic cytokine damage.

A protein of the present invention may also be useful for promoting or inhibiting differentiation of tissues described above from precursor tissues or cells; or for inhibiting the growth of tissues described above.

Alternatively, as described in more detail below, genes encoding these proteins or nucleic
25 acids regulating the expression of these proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

Assaying GENSET proteins or Fragments Thereof for Regulation of Reproductive Hormones or Cell Movement

The proteins encoded by the cDNAs of the invention or fragments thereof may also be
30 evaluated for their ability to regulate reproductive hormones, such as follicle stimulating hormone. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Vale *et al.*, *Endocrinology* 91:562-572, 1972; Ling *et al.*, *Nature* 321:779-782, 1986; Vale *et al.*, *Nature* 321:776-779, 1986; Mason *et al.*, *Nature* 318:659-663, 1985; Forage *et al.*, *Proc. Natl. Acad. Sci.*
35 *USA* 83:3091-3095, 1986. Chapter 6.12 (Measurement of Alpha and Beta Chemokines) Current Protocols in Immunology, J.E. Coligan *et al.* Eds. Greene Publishing Associates and Wiley-

Interscience ; Taub *et al. J. Clin. Invest.* 95:1370-1376, 1995; Lind *et al. APMIS* 103:140-146, 1995; Muller *et al. Eur. J. Immunol.* 25:1744-1748; Gruber *et al. J. of Immunol.* 152:5860-5867, 1994; Johnston *et al. J. of Immunol.* 153:1762-1768, 1994.

Those proteins which exhibit activity as reproductive hormones or regulators of cell movement may then be formulated as pharmaceuticals and used to treat clinical conditions in which regulation of reproductive hormones or cell movement are beneficial. For example, a protein of the present invention may also exhibit activin- or inhibin-related activities. Inhibins are characterized by their ability to inhibit the release of follicle stimulating hormone (FSH), while activins are characterized by their ability to stimulate the release of follicle stimulating hormone (FSH). Thus, a protein of the present invention, alone or in heterodimers with a member of the inhibin α family, may be useful as a contraceptive based on the ability of inhibins to decrease fertility in female mammals and decrease spermatogenesis in male mammals. Administration of sufficient amounts of other inhibins can induce infertility in these mammals. Alternatively, the protein of the invention, as a homodimer or as a heterodimer with other protein subunits of the inhibin-B group, may be useful as a fertility inducing therapeutic, based upon the ability of activin molecules in stimulating FSH release from cells of the anterior pituitary. See, for example, United States Patent 4,798,885, the disclosure of which is incorporated herein by reference. A protein of the invention may also be useful for advancement of the onset of fertility in sexually immature mammals, so as to increase the lifetime reproductive performance of domestic animals such as cows, sheep and pigs.

Alternatively, as described in more detail below, genes encoding these proteins or nucleic acids regulating the expression of these proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

Assaying GENSET proteins or Fragments Thereof for Chemotactic/Chemokinetic Activity

The proteins encoded by the cDNAs of the invention or fragments thereof may also be evaluated for chemotactic/chemokinetic activity. For example, a protein of the present invention may have chemotactic or chemokinetic activity (e.g., act as a chemokine) for mammalian cells, including, for example, monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells. Chemotactic and chemokinetic proteins can be used to mobilize or attract a desired cell population to a desired site of action. Chemotactic or chemokinetic proteins provide particular advantages in treatment of wounds and other trauma to tissues, as well as in treatment of localized infections. For example, attraction of lymphocytes, monocytes or neutrophils to tumors or sites of infection may result in improved immune responses against the tumor or infecting agent.

A protein or peptide has chemotactic activity for a particular cell population if it can stimulate, directly or indirectly, the directed orientation or movement of such cell population. Preferably, the protein or peptide has the ability to directly stimulate directed movement of cells.

Whether a particular protein has chemotactic activity for a population of cells can be readily determined by employing such protein or peptide in any known assay for cell chemotaxis.

The activity of a protein of the invention may, among other means, be measured by the following methods:

- 5 Assays for chemotactic activity (which will identify proteins that induce or prevent chemotaxis) consist of assays that measure the ability of a protein to induce the migration of cells across a membrane as well as the ability of a protein to induce the adhesion of one cell population to another cell population. Suitable assays for movement and adhesion include, without limitation, those described in: Current Protocols in Immunology, Ed by J.E. Coligan, A.M. Kruisbeek, D.H. Margulies, E.M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 6.12, Measurement of alpha and beta Chemokines 6.12.1-6.12.28; Taub *et al.* J. Clin. Invest. 95:1370-1376, 1995; Lind *et al.* APMIS 103:140-146, 1995; Mueller *et al.* Eur. J. Immunol. 25:1744-1748; Gruber *et al.* J. of Immunol. 152:5860-5867, 1994; Johnston *et al.* J. of Immunol, 153:1762-1768, 1994.

15 Assaying GENSET proteins or Fragments Thereof for Regulation of Blood Clotting

- The proteins encoded by the cDNAs of the invention or fragments thereof may also be evaluated for their effects on blood clotting. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Linet *et al.*, *J. Clin. Pharmacol.* 26:131-140, 1986; Burdick *et al.*, *Thrombosis Res.* 45:413-419, 1987; Humphrey *et al.*, *Fibrinolysis* 5:71-79 (1991); Schaub, *Prostaglandins* 35:467-474, 1988.

- Those proteins which are involved in the regulation of blood clotting may then be formulated as pharmaceuticals and used to treat clinical conditions in which regulation of blood clotting is beneficial. For example, a protein of the invention may also exhibit hemostatic or thrombolytic activity. As a result, such a protein is expected to be useful in treatment of various coagulations disorders (including hereditary disorders, such as hemophilias) or to enhance coagulation and other hemostatic events in treating wounds resulting from trauma, surgery or other causes. A protein of the invention may also be useful for dissolving or inhibiting formation of thromboses and for treatment and prevention of conditions resulting therefrom (such as, for example, infarction of cardiac and central nervous system vessels (e.g., stroke)). Alternatively, as described in more detail below, genes encoding these proteins or nucleic acids regulating the expression of these proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

Assaying GENSET proteins or Fragments Thereof for Involvement in Receptor/Ligand Interactions

- 35 The proteins encoded by the cDNAs or a fragment thereof may also be evaluated for their involvement in receptor/ligand interactions. Numerous assays for such involvement are familiar to

those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Chapter 7.28 (Measurement of Cellular Adhesion under Static Conditions 7.28.1-7.28.22) in Current Protocols in Immunology, J.E. Coligan *et al.* Eds. Greene Publishing Associates and Wiley-Interscience; Takai *et al.*, *Proc. Natl. Acad. Sci. USA* 84:6864-6868, 1987; Bierer *et al.*, *J. Exp. Med.* 168:1145-1156, 1988; Rosenstein *et al.*, *J. Exp. Med.* 169:149-160, 1989; Stoltenborg *et al.*, *J. Immunol. Methods* 175:59-68, 1994; Stitt *et al.*, *Cell* 80:661-670, 1995; Gyuris *et al.*, *Cell* 75:791-803, 1993.

For example, the proteins of the present invention may also demonstrate activity as receptors, receptor ligands or inhibitors or agonists of receptor/ligand interactions. Examples of such receptors and ligands include, without limitation, cytokine receptors and their ligands, receptor kinases and their ligands, receptor phosphatases and their ligands, receptors involved in cell-cell interactions and their ligands (including without limitation, cellular adhesion molecules (such as selectins, integrins and their ligands) and receptor/ligand pairs involved in antigen presentation, antigen recognition and development of cellular and humoral immune responses). Receptors and ligands are also useful for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. A protein of the present invention (including, without limitation, fragments of receptors and ligands) may themselves be useful as inhibitors of receptor/ligand interactions.

Assaying GENSET proteins or Fragments Thereof for Anti-Inflammatory Activity

The proteins encoded by the cDNAs or a fragment thereof may also be evaluated for anti-inflammatory activity. The anti-inflammatory activity may be achieved by providing a stimulus to cells involved in the inflammatory response, by inhibiting or promoting cell-cell interactions (such as, for example, cell adhesion), by inhibiting or promoting chemotaxis of cells involved in the inflammatory process, inhibiting or promoting cell extravasation, or by stimulating or suppressing production of other factors which more directly inhibit or promote an inflammatory response. Proteins exhibiting such activities can be used to treat inflammatory conditions including chronic or acute conditions), including without limitation inflammation associated with infection (such as septic shock, sepsis or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine-induced lung injury, inflammatory bowel disease, Crohn's disease or resulting from over production of cytokines such as TNF or IL-1. Proteins of the invention may also be useful to treat anaphylaxis and hypersensitivity to an antigenic substance or material.

Assaying GENSET proteins or Fragments Thereof for Tumor Inhibition Activity

The proteins encoded by the cDNAs of the invention or a fragment thereof may also be evaluated for tumor inhibition activity. In addition to the activities described above for immunological treatment or prevention of tumors, a protein of the invention may exhibit other anti-

tumor activities. A protein may inhibit tumor growth directly or indirectly (such as, for example, via ADCC). A protein may exhibit its tumor inhibitory activity by acting on tumor tissue or tumor precursor tissue, by inhibiting formation of tissues necessary to support tumor growth (such as, for example, by inhibiting angiogenesis), by causing production of other factors, agents or cell types
5 which inhibit tumor growth, or by suppressing, eliminating or inhibiting factors, agents or cell types which promote tumor growth.

A protein of the invention may also exhibit one or more of the following additional activities or effects: inhibiting the growth, infection or function of, or killing, infectious agents, including, without limitation, bacteria, viruses, fungi and other parasites; effecting (suppressing or
10 enhancing) bodily characteristics, including, without limitation, height, weight, hair color, eye color, skin, fat to lean ratio or other tissue pigmentation, or organ or body part size or shape (such as, for example, breast augmentation or diminution, change in bone form or shape); effecting biorhythms or circadian cycles or rhythms; effecting the fertility of male or female subjects; effecting the metabolism, catabolism, anabolism, processing, utilization, storage or elimination of
15 dietary fat, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional factors or component(s); effecting behavioral characteristics, including, without limitation, appetite, libido, stress, cognition (including cognitive disorders), depression (including depressive disorders) and violent behaviors; providing analgesic effects or other pain reducing effects; promoting differentiation and growth of embryonic stem cells in lineages other than hematopoietic lineages;
20 hormonal or endocrine activity; in the case of enzymes, correcting deficiencies of the enzyme and treating deficiency-related diseases; treatment of hyperproliferative disorders (such as, for example, psoriasis); immunoglobulin-like activity (such as, for example, the ability to bind antigens or complement); and the ability to act as an antigen in a vaccine composition to raise an immune response against such protein or another material or entity which is cross-reactive with such protein.

25

References

- Abbondanzo *et al.*, (1993), *Meth. Enzymol.*, Academic Press, New York, pp 803-823
Altschul *et al.*, (1990), *J. Mol. Biol.* 215(3):403-410
Altschul *et al.*, (1993), *Nature Genetics* 3:266-272
Altschul *et al.*, (1997), *Nuc. Acids Res.* 25:3389-3402
30 Ames *et al.*, (1995), *J. Immunol. Meth.* 184:177-186.
Anton and Graham, (1995), *J. Virol.*, 69: 4600-4606
Araki *et al.*, (1995) *Proc. Natl. Acad. Sci. U S A.* 92(1):160-4.
Ashkenazi *et al.*, (1991), *Proc. Natl. Acad. Sci. U S A* 88:10535-10539.
Aszodi *et al.*, (1997) *Proteins: Structure, Function, and Genetics*, Supplement 1:38-42
35 Attwood *et al.*, (1996) *Nucleic Acids Res.* 24(1):182-8.
Attwood *et al.*, (2000) *Nucleic Acids Res.* 28(1):225-7

- Bartunek *et al.*, (1996), *Cytokine*. 8(1):14-20.
- Bateman *et al.*, (2000) *Nucleic Acids Res.* 28(1):263-6
- Baubonis (1993) *Nucleic Acids Res.* 21(9):2025-9.
- Beaucage *et al.*, (1981) *Tetrahedron Lett*, 22: 1859-1862
- 5 Benham *et al.* (1989) *Genomics* 4:509-517,
- Better *et al.*, (1988), *Science*. 240:1041-1043.
- Bhagwat *et al.* (1999) *J. Biol. Chem.* 274:24014-22
- Bittle *et al.*, (1985), *Viol.* 66:2347-2354.
- Bowie *et al.*, (1994), *Science*. 247:1306-1310.
- 10 Bradley (1987), Production and analysis of chimaeric mice. *In*: E.J. Robertson (Ed.),
Teratocarcinomas and embryonic stem cells: A practical approach. IRL Press, Oxford, pp.113.
- Bram *et al.*, (1993), *Mol. Cell Biol.*, 13 : 4760-4769
- Brinkman *et al.*, (1995) *J. Immunol Methods*. 182:41-50.
- Brown *et al.*, (1979) *Meth. Enzymol.* 68:109-151
- 15 Brutlag *et al.* (1990) *Comp. App. Biosci.* 6:237-245
- Bucher and Bairoch (1994) Proceedings 2nd International Conference on Intelligent
Systems for Molecular Biology. Altman *et al*, Eds., pp53-61, AAAIPress, Menlo Park.
- Burton *et al.* (1994), *Adv. Immunol.* 57:191-280
- Bush *et al.*, (1997), *J. Chromatogr.*, 777 : 311-328.
- 20 Butt and Karathanasis (1995) *Gene Expr.* 4(6):319-36.
- Carlson *et al.*, (1997), *J. Biol. Chem.* 272(17):11295-11301.
- Chai *et al.*, (1993) *Biotechnol. Appl. Biochem.* 18:259-273.
- Chang *et al.*, (1993) *Gene* 127:95-8
- Chee *et al.*, (1996) *Science*. 274:610-614.
- 25 Chen *et al.* (1987) *Mol. Cell. Biol.* 7:2745-2752.
- Chen *et al.*, (1998), *Cancer Res.* 58(16):3668-3678.
- Cherif *et al.*, (1990) *Proc. Natl. Acad. Sci. U.S.A.*, 87:6639-6643
- Cho *et al.*, (1998), *Proc. Natl. Acad. Sci. USA*, 95(7) : 3752-3757.
- Chou, (1989), *Mol. Endocrinol.* 3: 1511-1514.
- 30 Chow *et al.*, (1985), *Proc. Natl. Acad. Sci. USA*. 82:910-914.
- Christophe *et al.* (2000) *Cell Signal* 12(5):337-41
- Ciminale *et al.* (1999) *Oncogene* 18:4505-14
- Cleland *et al.*, (1993), *Crit. Rev. Therapeutic Drug Carrier Systems.* 10:307-377.
- Coles *et al.*, (1998) *Hum Mol Genet* 7:791-800
- 35 Compton (1991) *Nature*. 350(6313):91-92.
- Corpet *et al.* (2000) *Nucleic Acids Res.* 28(1):267-9
- Cox *et al.*, (1990) *Science* 250:245-250

- Creighton (1983), *Proteins: Structures and Molecular Principles*, W.H. Freeman & Co. 2nd Ed., T. E., New York
- Creighton, (1993) , *Posttranslational Covalent Modification of Proteins*, W.H. Freeman and Company, New York B.C. Johnson, Ed., Academic Press, New York 1-12
- 5 Cunningham *et al.* (1989), *Science* 244:1081-1085.
- Davis *et al.*, (1986) *Basic Methods in Molecular Biology*, ed., Elsevier Press, NY,
- Decker and Parker, (1995) *Curr. Opin. Cell. Biol.* 7(3) :368-92
- Dempster *et al.*, (1977) *Stat. Soc.*, 39B:1-38.
- Deng *et al.*, (1998) *Blood*. 92(6):1981-1988.
- 10 Dent and Latchman (1993) *The DNA mobility shift assay*. In: *Transcription Factors: A Practical Approach* (Latchman DS, ed.) pp1-26. Oxford: IRL Press
- Derrigo *et al.*, (2000) *Int. J. Mol. Med.* 5(2) :111-23
- Eckner *et al.*, (1991) *EMBO J.* 10:3513-3522.
- Edwards and Leatherbarrow, (1997) *Analytical Biochemistry*, 246, 1-6
- 15 Engvall, (1980) *Meth. Enzymol.* 70:419
- Erlich, (1992) *PCR Technology; Principles and Applications for DNA Amplification*. W.H. Freeman and Co., New York
- Essl *et al.*, (1999) *FEBS Lett.* 453:169-73.
- Feldman and Steg, (1996), *Medecine/Sciences*, 12:47-55
- 20 Felgner (1996) *Hum Gene Ther.* 7(15):1791-3.
- Felici, (1991), *J. Mol. Biol.*, 222:301-310
- Fell *et al.*, (1991), *J. Immunol.* 146:2446-2452.
- Fields and Song, (1989), *Nature*, 340: 245-246
- Fisher, (1980) Chap. 42 in: *Manual of Clinical Immunology*, 2d Ed. (Rose and Friedman,
- 25 Eds.) *Amer. Soc. For Microbiol.*, Washington, D.C.
- Flotte *et al.*, (1992) *Am. J. Respir. Cell Mol. Biol.* 7:349-356.
- Fodor *et al.*, (1991) *Science* 251:767-777.
- Foster *et al.*, (1996) *Genomics* 33:185-192
- Fountoulakis *et al.*, (1995) *Biochem.* 270:3958-3964.
- 30 Fraley *et al.*, (1979) *Proc. Natl. Acad. Sci. USA.* 76:3348-3352.
- Frazer *et al.*, (1992) *Genomics* 14:574-584
- Fried and Crothers, (1981) *Nucleic Acids Res.* 9:6505-6525
- Fromont-Racine *et al.*, (1997), *Nature Genetics*, 16(3) : 277-282.
- Fry *et al.*, (1992) *Biotechniques*, 13: 124-131
- 35 Fudenberg, (1980) Chap. 26 in: *Basic 503 Clinical Immunology*, 3rd Ed. Lange, Los Altos, California
- Fuller S. A. *et al.* (1996) *Immunology in Current Protocols in Molecular Biology*,

- Furth P.A. *et al.* (1994) *Proc. Natl. Acad. Sci USA*. 91:9302-9306.
- Garner and Revzin, (1981) *Nucleic Acids Res* 9:3047-3060
- Gentz *et al.*, (1989) *Proc Natl Acad Sci U S A*. 86(3):821-4.
- Geysen *et al.*, (1984), *Proc. Natl. Acad. Sci. U.S.A.* 81:3998-4002.
- 5 Ghosh and Bacchawat, (1991), Targeting of liposomes to hepatocytes, IN: Liver Diseases, Targeted diagnosis and therapy using specific receptors and ligands. Eds., Marcel Dekker, New York, pp. 87-104.
- Gillies *et al.*, (1989), *J. Immunol Methods*. 125:191-202.
- Gillies *et al.*, (1992), *Proc Natl Acad Sci U S A* 89:1428-1432.
- 10 Glaser *et al.* (1998) *Plant Mol Biol* 38:311-38;
- Gonnet *et al.*, (1992), *Science* 256:1443-1445
- Gopal (1985) *Mol. Cell. Biol.*, 5:1188-1190.
- Gossen *et al.*, (1992) *Proc. Natl. Acad. Sci. USA*. 89:5547-5551.
- Gossen *et al.*, (1995) *Science*. 268:1766-1769.
- 15 Graham *et al.*, (1973) *Virology* 52:456-457.
- Green *et al.*, (1986) *Ann. Rev. Biochem.* 55:569-597
- Greenspan and Bona (1989), *FASEB J.* 7(5):437-444.
- Griffais *et al.*, (1991) *Nucleic Acids Res.* 19: 3887-3891
- Griffin *et al.*, (1989) *Science* 245:967-971
- 20 Gu H. *et al.*, (1993) *Cell* 73:1155-1164.
- Gu H. *et al.*, (1994) *Science* 265:103-106.
- Guatelli *et al.*, (1990) *Proc. Natl. Acad. Sci. USA*. 35:273-286.
- Gyuris *et al.*, (1993) *Cell* 75:791-803
- Hames and Higgins (1985) *Nucleic Acid Hybridization: A Practical Approach*. Hames and
- 25 Higgins Ed., IRL Press, Oxford.
- Hammerling (1981), *Monoclonal Antibodies and T-Cell Hybridomas*, Elsevier, N.Y. 563-681.
- Hansson *et al.*, (1999), *J. Mol. Biol.* 287:265-276.
- Haravama (1998), *Trends Biotechnol.* 16(2): 76-82.
- 30 Harland *et al.*, (1985) *J. Cell. Biol.* 101:1094-1095.
- Harlow and Lane, (1988) *Antibodies A Laboratory Manual*. Cold Spring Harbor Laboratory. pp. 53-242
- Harper *et al.*, (1993), *Cell*, 75 : 805-816
- Harrop *et al.*, (1998), *J. Immunol.* 161(4):1786-1794.
- 35 Haynes *et al.*, (1996) *J Biotechnol.* 44(1-3):37-42.
- Henikoff and Henikoff, (1993), *Proteins* 17:49-61
- Henikoff *et al.*, (2000) *Electrophoresis* 21(9):1700-6

- Henikoff *et al.*, (2000) *Nucleic Acids Res.* 28(1):228-30
- Herrman and Neupert, (2000) *Curr. Opinion Microbiol.* 3:210-4
- Higgins *et al.*, (1996), *Meth. Enzymol.* 266:383-402
- Hillier and Green (1991) *PCR Methods Appl.*, 1: 124-8.
- 5 Hoess *et al.*, (1986) *Nucleic Acids Res.* 14:2287-2300.
- Hofmann *et al.*, (1999) *Nucl. Acids Res.* 27:215-219.;
- Holm and Sander (1996) *Nucleic Acids Res.* 24(1):206-9
- Holm and Sander (1997) *Nucleic Acids Res.* 25(1):231-4
- Holm and Sander (1999) *Nucleic Acids Res.* 27(1):244-7
- 10 Hoppe *et al.*, (1994), *FEBS Letters.* 344:191.
- Houghten (1985), *Proc. Natl. Acad. Sci. USA* 82:5131-5135.
- Huang *et al.*, (1996) *Cancer Res* 56(5):1137-1141.
- Hunkapiller *et al.*, (1984) *Nature.* 310(5973):105-11.
- Huston *et al.*, (1991), *Meth. Enzymol.* 203:46_88.
- 15 Huygen *et al.*, (1996) *Nature Medicine.* 2(8):893-898.
- Izant and Weintraub, (1984) *Cell* 36(4):1007-15
- Jameson and Wolf, (1988), *Comp. Appl. Biosci.* 4:181-186
- Julan *et al.*, (1992) *J. Gen. Virol.* 73:3251-3255.
- Kanegae *et al.*, (1995) *Nucl. Acids Res.* 23:3816-3821.
- 20 Karlin and Altschul, (1990), *Proc. Natl. Acad. Sci. USA* 87:2267-2268
- Kettleborough *et al.*, (1994), *Eur. L Immunol.* 24:952-958.
- Kim U-J. *et al.*, (1996) *Genomics* 34:213-218.
- Klein *et al.*, (1987) *Nature.* 327:70-73.
- Kohler and Milstein, (1975) *Nature* 256:495
- 25 Koller *et al.*; (1992) *Annu. Rev. Immunol.* 10:705-730.
- Kostelny *et al.*, (1992), *J. Immunol.* 148:1547-1553.
- Landschulz *et al.*, (1988), *Science.* 240:1759.
- Ledbetter *et al.*, (1990) *Genomics* 6:475-481
- Lenhard *et al.*, (1996) *Gene.* 169:187-190.
- 30 Levy *et al.*, (1996) *Gene Ther.* 3(3):201-11.
- Lewin, (1989), *Proc. Natl. Acad. Sci. USA* 86:9832-8935.
- Liautard *et al.*, (1997), *Cytokine.* 9(4):233-241.
- Linton *et al.*, (1993) *J. Clin. Invest.* 92:3029-3037.
- Liu *et al.*, (1994) *Proc. Natl. Acad. Sci. USA.* 91: 4528-4262.
- 35 Lo Conte *et al.*, (2000) *Nucleic Acids Res.* 28(1):257-9.
- Lockhart *et al.*, (1996) *Nature Biotechnology* 14: 1675-1680
- Lorenzo and Blasco (1998) *Biotechniques.* 24(2):308-313.

- Lucas (1994), In : Development and Clinical Uses of Haemophilus b Conjugate;
 Luetterforst et al., (1999) *J. Cell. Biol.* 145:1443-59
 Makrides, (1999) *Protein Expr. Purif.* 17(2) :183-202
 Malik et al., (1992), *Exp. Hematol.* 20:1028-1035.
 5 Mansour et al., (1988) *Nature.* 336:348-352.
 Marshall et al., (1994) *PCR Methods and Applications.* 4:80-84.
 Maurer et al., (1999) *Mol Membr Biol.* 16(1):129-40.
 McCormick et al., (1994) *Genet. Anal. Tech. Appl.* 11:158-164.
 McLaughlin et al., (1996) *Am. J. Hum. Genet.* 59:561-569.
 10 Miller and Whelan, (1997) *Hum Gene Ther.* 8(7):803-15.
 Muller et al., (1998), *Structure.* 6(9):1153-1167.
 Mullinax et al., (1992), *BioTechniques.* 12(6):864-869.
 Munro, (1998) *Trends Cell Biol.* 8:11-15
 Munro and Pelham (1987) *Cell* 48(5):899-907
 15 Murphy (1997) *Trends Biotechnol.* 15:326-30
 Murvai et al., (2000) *Nucleic Acids Res.* 28(1):260-2
 Murzin et al., (1995) *J Mol Biol.* 247(4):536-40
 Muzyczka et al., (1992) *Curr. Topics in Micro. and Immunol.* 158:97-129.
 Nada et al., (1993) *Cell* 73:1125-1135.
 20 Nagaraja et al, (1997. *Genome Research* 7:210-222
 Nagy et al., (1993), *Proc. Natl. Acad. Sci. USA* 90: 8424-8428.
 Nakai and Horton, (1999) *Trends Biochem. Sci.*, 24:34-36
 Nakai and Kanehisa (1992) *Genomics* 14, 897-911
 Naramura et al., (1994), *Immunol. Lett.* 39:91-99.
 25 Narang et al., (1979), *Methods Enzymol* 68:90-98
 Neda et al., (1991) *J. Biol. Chem.* 266:14143-14146.
 Nevill-Manning et al., (1998) *Proc. Natl. Acad. Sci. U S A.* 95, 5865-5871
 Nicolau et al., (1982) *Biochim. Biophys. Acta.* 721:185-190.
 Nicolau et al., (1987), *Meth. Enzymol.*, 149:157-76.
 30 Nissinoff, (1991), *J. Immunol.* 147(8): 2429-2438.
 O'Reilly et al., (1992) *Baculovirus Expression Vectors: A Laboratory Manual.* W. H.
 Freeman and Co., New York.
 Obermayr et al., (1996) *Eur. J. Hum. Genet.* 4:242-245
 Ohno et al., (1994) *Science.* 265:781-784.
 35 Oi et al., (1986), *BioTechniques* 4:214.
 Oldenburg et al., (1992), *Proc. Natl. Acad. Sci. USA* 89:5393-5397.
 Orenge et al., (1997) *Structure.* 5(8):1093-108

- Ouchterlony *et al.*, (1973) Chap. 19 in: Handbook of Experimental Immunology D. Wier (ed) Blackwell
- Padlan, (1991), *Molec. Immunol.* 28(4/5):489-498.
- Parmley and Smith, (1988) *Gene* 73:305-318
- 5 Patten, *et al.* (1997), *Curr Opin Biotechnol.* 8:724-733.
- Pearl *et al.*, (2000) *Biochem Soc Trans.* 28(2):269-75
- Pearson and Lipman, (1988), *Proc. Natl. Acad. Sci. USA* 85(8):2444-2448
- Pease and William, (1990), *Exp. Cell. Res.* 190: 209-211.
- Pelham (1990) *Trends Biochem Sci* 12:483-6
- 10 Persic *et al.*, (1997), *Gene.* 1879-81
- Pesole *et al.*, (2000) *Nucleic Acids Res*, 28(1):193-196
- Peterson *et al.*, (1993), *Proc. Natl. Acad. Sci. USA*, 90 : 7593-7597.
- Picetti and Borrelli, (2000) *Exp Cell Res* 255:258-69
- Pidoux and Armstrong (1992) *EMBO J* Apr;11(4):1583-91
- 15 Pietu *et al.*, (1996) *Genome Research* 6:492-503
- Pinckard *et al.*, (1967), *Clin. Exp. Immunol* 2:331-340.
- Pitard *et al.*, (1997), *J. Immunol. Methods.* 205(2):177-190.
- Pongor *et al.* (1993) *Protein Eng.* 6(4):391-5
- Potter *et al.*, (1984) *Proc. Natl. Acad. Sci. U.S.A.* 81(22):7161-7165.
- 20 Prat *et al.*, (1998), *J. Cell. Sci.* 111(Pt2):237-247.
- Raeymaekers *et al.*, (1995) *Genomics* 29:170-178
- Ramunsen *et al.*, (1997), *Electrophoresis*, 18 : 588-598.
- Rattan *et al.*, (1992) *Ann NY Acad Sci* 663:48-62
- Reid *et al.*, (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87:4299-4303.
- 25 Robbins *et al.*, (1987), *Diabetes.* 36:838-845.
- Robertson, (1987), Embryo-derived stem cell lines. In: E.J. Robertson Ed. Teratocarcinomas and embryonic stem cells: a practical approach. IRL Press, Oxford, pp. 71.
- Roguska *et al.*, (1994), *Proc. Natl. Acad. Sci. U.S.A.* 91:969-973.
- Ron *et al.*, (1993), *Biol Chem.*, 268 2984-2988.
- 30 Rose *et al.*, (1980) Chap. 12 in: *Methods in Immunodiagnosis*, 2d Ed. John Wiley 503 Sons, New York
- Rossi *et al.*, (1991) *Pharmacol. Ther.* 50:245-254,
- Roth *et al.*, (1996) *Nature Medicine.* 2(9):985-991.
- Roux *et al.*, (1989) *Proc. Natl. Acad. Sci. U.S.A.* 86:9079-9083.
- 35 Sambrook *et al.*, (1989) *Molecular Cloning: A Laboratory Manual.* 2ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
- Samson *et al.*, (1996) *Nature*, 382(6593):722-725.

- Samulski *et al.*, (1989) J. Virol. 63:3822-3828.
- Sanchez-Pescador (1988) J. Clin. Microbiol. 26(10):1934-1938.
- Sander and Schneider (1991) Proteins. 9(1):56-68.)
- Sauer *et al.*, (1988) Proc. Natl. Acad. Sci. U.S.A. 85:5166-5170.
- 5 Sawai *et al.*, (1995), AJRI 34:26-34.
- Schedl *et al.*, (1993b), Nucleic Acids Res., 21: 4783-4787.
- Schedl *et al.*, (1993a), Nature, 362: 258-261.
- Schena *et al.* (1995) Science 270:467-470
- Schena *et al.*, (1996), Proc Natl Acad Sci U S A, 93(20):10614-10619.
- 10 Schuler *et al.*, (1996) Science 274:540-546
- Schultz *et al.*, (1998) Proc Natl Acad Sci U S A 95, 5857-5864
- Schwartz and Dayhoff, (1978), eds., Matrices for Detecting Distance Relationships: Atlas of Protein Sequence and Structure, Washington: National Biomedical Research Foundation
- Sczakiel *et al.*, (1995) Trends Microbiol. 3(6):213-217.
- 15 Seifter *et al.*, (1990) Meth Enzymol 182:626-646
- Shay *et al.*, (1991), Biochem. Biophys. Acta, 1072: 1-7.
- Shizuya *et al.*, (1992) Proc. Natl. Acad. Sci. U.S.A. 89:8794-8797.
- Shu *et al.*, (1993), Proc. Natl. Acad. Sci. U.S.A. 90:7995-7999.
- Skerra *et al.*, (1988), Science 240:1038-1040.
- 20 Smith and Johnson (1988) Gene. 67(1):31-40.
- Smith *et al.*, (1983) Mol. Cell. Biol. 3:2156-2165.
- Smith *et al.*, (1996) Antiviral Res. 32(2):99-115.
- Sonnhammer and Kahn D (1994) Protein Sci. 3(3):482-92
- Sonnhammer *et al.*, (1997) Proteins. 28(3):405-20
- 25 Sosnowski, *et al.*, (1997) Proc Natl Acad Sci U S A 94:1119-1123
- Sowdhamini *et al.*, (1997) Protein Engineering 10:207, 215
- Sternberg (1994) Mamm. Genome. 5:397-404.
- Sternberg (1992) Trends Genet. 8:1-16.
- Sternberg *et al.*, (1999) Curr Opin Struct Biol. 9(3):368-73.
- 30 Stone *et al.*, (2000) J Endocrinol. 164(2):103-18.
- Stryer, (1995) Biochemistry, 4th edition
- Studnicka *et al.*, (1994), Protein Engineering. 7(6):805-814.
- Sutcliffe *et al.*, (1983), Science. 219:660-666.
- Szabo *et al.*, (1995) Curr Opin Struct Biol 5, 699-705
- 35 Tacson *et al.*, (1996) Nature Medicine. 2(8):888-892.
- Taryman *et al.*, (1995), Neuron. 14(4):755-762.
- Tatusov *et al.*, (1997) Science, 278, 631 :637

- Tatusov *et al.*, (2000) *Nucleic Acids Res.* 28(1):33-6.)
Te Riele *et al.*, (1990) *Nature.* 348:649-651.
Thomas *et al.*, (1986) *Cell.* 44:419-428.
Thomas *et al.*, (1987) *Cell.* 51:503-512.
5 Thompson *et al.*, (1994), *Nucleic Acids Res.* 22(2):4673-4680
Traunecker *et al.*, (1988), *Nature.* 331:84-86.
Tur-Kaspa *et al.*, (1986) *Mol. Cell. Biol.* 6:716-718.
Tutt *et al.*, (1991), *J. Immunol.* 147:60-69.
Ugur and Jones (2000) *Mol Cell Biol* 11:1432-32
10 Urdea (1988) *Nucleic Acids Research.* 11:4937-4957.
Urdea *et al.*, (1991) *Nucleic Acids Symp. Ser.* 24:197-200.
Vaitukaitis *et al.*, (1971) *J. Clin. Endocrinol. Metab.* 33:988-991
Valadon *et al.*, (1996), *J. Mol. Biol.*, 261:11-22.
Van der Lugt *et al.*, (1991) *Gene.* 105:263-267.
15 Vil *et al.*, (1992) *Proc Natl Acad Sci U S* 89:11337-11341.
Vlasak *et al.*, (1983) *Eur. J. Biochem.* 135:123-126.
Wabiko *et al.*, (1986) *DNA*.5(4):305-314.
Wagner *et al.*, (1996) *Nat Biotechnol.* 14(7):840-4.
Walker *et al.*, (1996) *Clin. Chem.* 42:9-13.
20 Wang *et al.*, (1997), *Chromatographia*, 44 : 205-208.
Ward and Moss, (2000) *J. Virol.* 74:3771-80
Warrington *et al.*, (1991) *Genomics* 11:701-708
Watson and Pessin, (2000) *J Biol Chem* 275:1261-8
Westerink, (1995), *Proc. Natl. Acad. Sci USA.*, 92:4021-4025
25 White (1997) B.A. Ed. in *Methods in Molecular Biology* 67: Humana Press, Totowa
White *et al.* (1997) *Genomics.* 12:301-306.
Wilson *et al.*, (1984) *Cell.* 37(3):767-78.
Wong *et al.*, (1980) *Gene.* 10:87-94.
Wood *et al.*, (1985) *Proc. Natl. Acad. Sci. USA* 82(6):1585-1588
30 Wood *et al.*, (1993), *Proc. Natl. Acad. Sci. USA*, 90: 4582-4585.
Wu and Ataai, (2000) *Curr Opin Biotechnol.* 11(2):205-8.
Wu and Wu, (1987) *J. Biol. Chem.* 262:4429-4432.
Wu and Wu, (1988) *Biochemistry.* 27:887-892.
Yagi T. *et al.*, (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87:9918-9922.
35 Yona *et al.*, (1999) *Proteins.* 37(3):360-78
Yoon *et al.*, (1998), *J. Immunol.* 160(7):3170-3179.
Zhan *et al.*, (1998) *Cancer Immunol Immunother* 46:55-60

Zheng, X.X. *et al.* (1995), J. Immunol. 154:5590-5600.

Zhu *et al.*, (1998), Cancer Res. 58(15):3209-3214.

Zou *et al.*, (1994) Curr. Biol. 4:1099-1103.

Throughout this application, various publications, patents and published patent applications
5 are cited. The disclosures of these publications, patents and published patent specification
referenced in this application are hereby incorporated by reference into the present disclosure to
more fully describe the state of the art to which this invention pertains.

Table I

Seq Id No	Internal designation	Type	Vector
1	119-003-4-0-C2-CS	DNA	pBluescriptII SK-
2	105-016-1-0-D3-CS	DNA	pBluescriptII SK-
3	105-016-3-0-G10-CS	DNA	pBluescriptII SK-
4	105-026-1-0-A5-CS	DNA	pBluescriptII SK-
5	105-031-1-0-A11-CS	DNA	pBluescriptII SK-
6	105-031-2-0-D3-CS	DNA	pBluescriptII SK-
7	105-035-2-0-C6-CS	DNA	pBluescriptII SK-
8	105-037-2-0-H11-CS	DNA	pBluescriptII SK-
9	105-053-4-0-E8-CS	DNA	pBluescriptII SK-
10	105-074-3-0-H10-CS	DNA	pBluescriptII SK-
11	105-089-3-0-G10-CS	DNA	pBluescriptII SK-
12	105-095-2-0-G11-CS	DNA	pBluescriptII SK-
13	106-006-1-0-E3-CS	DNA	pBluescriptII SK-
14	106-037-1-0-E9-CS.cor	DNA	pBluescriptII SK-
15	106-037-1-0-E9-CS.fr	DNA	pBluescriptII SK-
16	106-043-4-0-H3-CS	DNA	pBluescriptII SK-
17	110-007-1-0-C7-CS	DNA	pBluescriptII SK-
18	114-016-1-0-H8-CS	DNA	pBluescriptII SK-
19	116-004-3-0-A6-CS	DNA	pBluescriptII SK-
20	116-054-3-0-E6-CS	DNA	pBluescriptII SK-
21	116-055-1-0-A3-CS	DNA	pBluescriptII SK-
22	116-055-2-0-F7-CS	DNA	pBluescriptII SK-
23	116-088-4-0-A9-CS	DNA	pBluescriptII SK-
24	116-091-1-0-D9-CS	DNA	pBluescriptII SK-
25	116-110-2-0-F4-CS	DNA	pBluescriptII SK-
26	116-111-1-0-H9-CS	DNA	pBluescriptII SK-
27	116-111-4-0-B3-CS	DNA	pBluescriptII SK-
28	116-115-2-0-F8-CS	DNA	pBluescriptII SK-
29	116-119-3-0-H5-CS	DNA	pBluescriptII SK-
30	117-001-5-0-G3-CS	DNA	pBluescriptII SK-
31	145-25-3-0-B4-CS.cor	DNA	pBluescriptII SK-
32	145-25-3-0-B4-CS.fr	DNA	pBluescriptII SK-
33	145-56-3-0-D5-CS	DNA	pBluescriptII SK-
34	145-59-2-0-A7-CS	DNA	pBluescriptII SK-
35	157-15-4-0-B11-CS	DNA	pBluescriptII SK-
36	160-103-1-0-F11-CS	DNA	pBluescriptII SK-
37	160-37-2-0-H7-CS	DNA	pBluescriptII SK-
38	160-58-3-0-H3-CS	DNA	pBluescriptII SK-
39	160-75-4-0-A9-CS	DNA	pBluescriptII SK-
40	174-10-2-0-F8-CS	DNA	pPT
41	174-33-3-0-F6-CS	DNA	pPT
42	174-38-1-0-B6-CS	DNA	pPT
43	174-38-3-0-C9-CS	DNA	pPT
44	174-39-2-0-A3-CS	DNA	pPT
45	174-41-1-0-A6-CS	DNA	pPT
46	174-5-3-0-H7-CS	DNA	pPT
47	174-7-4-0-H1-CS	DNA	pPT
48	175-1-3-0-E5-CS.cor	DNA	pPT

49	175-1-3-0-E5-CS.fr	DNA	pPT
50	180-19-4-0-F4-CS	DNA	pBluescriptII SK-
51	181-10-1-0-D10-CS	DNA	pBluescriptII SK-
52	181-16-1-0-G7-CS	DNA	pBluescriptII SK-
53	181-16-2-0-A7-CS	DNA	pBluescriptII SK-
54	181-20-3-0-B5-CS	DNA	pBluescriptII SK-
55	181-3-3-0-B8-CS	DNA	pBluescriptII SK-
56	181-3-3-0-C9-CS	DNA	pBluescriptII SK-
57	182-1-2-0-D12-CS	DNA	pBluescriptII SK-
58	184-1-4-0-C11-CS	DNA	pBluescriptII SK-
59	184-4-1-0-A11-CS	DNA	pBluescriptII SK-
60	187-12-4-0-A8-CS	DNA	pBluescriptII SK-
61	187-2-2-0-A3-CS	DNA	pBluescriptII SK-
62	187-31-0-0-f12-CS	DNA	pBluescriptII SK-
63	187-34-0-0-l12-CS	DNA	pBluescriptII SK-
64	187-37-0-0-c10-CS	DNA	pBluescriptII SK-
65	187-38-0-0-l10-CS	DNA	pBluescriptII SK-
66	187-39-0-0-k12-CS	DNA	pBluescriptII SK-
67	187-41-0-0-i21-CS	DNA	pBluescriptII SK-
68	188-11-1-0-B3-CS	DNA	pBluescriptII SK-
69	188-18-4-0-A9-CS	DNA	pBluescriptII SK-
70	188-28-4-0-B12-CS.cor	DNA	pBluescriptII SK-
71	188-28-4-0-B12-CS.fr	DNA	pBluescriptII SK-
72	188-28-4-0-D4-CS	DNA	pBluescriptII SK-
73	188-41-1-0-B8-CS.cor	DNA	pBluescriptII SK-
74	188-41-1-0-B8-CS.fr	DNA	pBluescriptII SK-
75	188-45-1-0-D9-CS	DNA	pBluescriptII SK-
76	188-9-2-0-E1-CS	DNA	pBluescriptII SK-
77	105-079-3-0-A11-CS	DNA	pBluescriptII SK-
78	105-092-1-0-H7-CS	DNA	pBluescriptII SK-
79	105-141-4-0-H9-CS	DNA	pBluescriptII SK-
80	109-013-1-0-B9-CS	DNA	pBluescriptII SK-
81	110-008-4-0-D9-CS	DNA	pBluescriptII SK-
82	114-001-3-0-A2-CS	DNA	pBluescriptII SK-
83	114-028-2-0-C1-CS	DNA	pBluescriptII SK-
84	114-032-1-0-H10-CS	DNA	pBluescriptII SK-
85	114-043-2-0-A10-CS	DNA	pBluescriptII SK-
86	114-044-1-0-C5-CS	DNA	pBluescriptII SK-
87	116-003-3-0-D10-CS	DNA	pBluescriptII SK-
88	116-003-3-0-G12-CS	DNA	pBluescriptII SK-
89	116-011-2-0-F11-CS	DNA	pBluescriptII SK-
90	116-033-3-0-E4-CS	DNA	pBluescriptII SK-
91	116-041-4-0-B6-CS	DNA	pBluescriptII SK-
92	116-044-2-0-C4-CS	DNA	pBluescriptII SK-
93	116-075-1-0-E6-CS	DNA	pBluescriptII SK-
94	116-094-4-0-G5-CS	DNA	pBluescriptII SK-
95	117-005-3-0-F2-CS	DNA	pBluescriptII SK-
96	121-007-3-0-D9-CS	DNA	pBluescriptII SK-
97	145-91-3-0-D10-CS	DNA	pBluescriptII SK-
98	157-17-1-0-F4-CS	DNA	pBluescriptII SK-
99	160-11-3-0-G8-CS	DNA	pBluescriptII SK-
100	160-24-1-0-F12-CS	DNA	pBluescriptII SK-

101	160-24-2-0-E9-CS	DNA	pBluescriptII SK-
102	160-25-4-0-D2-CS	DNA	pBluescriptII SK-
103	160-31-3-0-A11-CS	DNA	pBluescriptII SK-
104	160-32-1-0-F6-CS	DNA	pBluescriptII SK-
105	160-37-1-0-A3-CS	DNA	pBluescriptII SK-
106	160-40-3-0-E9-CS	DNA	pBluescriptII SK-
107	160-58-3-0-E4-CS	DNA	pBluescriptII SK-
108	160-85-3-0-D4-CS	DNA	pBluescriptII SK-
109	160-95-3-0-A11-CS	DNA	pBluescriptII SK-
110	162-10-4-0-F9-CS.cor	DNA	pBluescriptII SK-
111	162-10-4-0-F9-CS.fr	DNA	pBluescriptII SK-
112	174-13-2-0-E4-CS	DNA	pPT
113	174-46-2-0-B11-CS	DNA	pPT
114	179-8-2-0-A6-CS	DNA	pBluescriptII SK-
115	180-22-3-0-B6-CS	DNA	pBluescriptII SK-
116	181-13-1-0-F7-CS	DNA	pBluescriptII SK-
117	181-15-4-0-F7-CS	DNA	pBluescriptII SK-
118	181-20-1-0-G7-CS	DNA	pBluescriptII SK-
119	184-15-3-0-D1-CS	DNA	pBluescriptII SK-
120	187-12-2-0-G11-CS	DNA	pBluescriptII SK-
121	187-2-2-0-A12-CS	DNA	pBluescriptII SK-
122	187-30-0-0-k23-CS	DNA	pBluescriptII SK-
123	187-36-0-0-e19-CS	DNA	pBluescriptII SK-
124	187-38-0-0-d22-CS	DNA	pBluescriptII SK-
125	187-39-0-0-b9-CS	DNA	pBluescriptII SK-
126	187-39-0-0-g6-CS	DNA	pBluescriptII SK-
127	187-45-0-0-l18-CS	DNA	pBluescriptII SK-
128	187-45-0-0-m21-CS	DNA	pBluescriptII SK-
129	187-45-0-0-n8-CS	DNA	pBluescriptII SK-
130	187-46-0-0-f23-CS	DNA	pBluescriptII SK-
131	187-5-1-0-A12-CS	DNA	pBluescriptII SK-
132	187-5-1-0-F6-CS	DNA	pBluescriptII SK-
133	187-5-2-0-B2-CS	DNA	pBluescriptII SK-
134	187-5-3-0-D5-CS	DNA	pBluescriptII SK-
135	187-51-0-0-f9-CS	DNA	pBluescriptII SK-
136	187-6-1-0-B9-CS	DNA	pBluescriptII SK-
137	187-6-4-0-C10-CS	DNA	pBluescriptII SK-
138	188-19-2-0-C8-CS	DNA	pBluescriptII SK-
139	188-22-4-0-G6-CS	DNA	pBluescriptII SK-
140	188-28-4-0-D11-CS	DNA	pBluescriptII SK-
141	188-29-1-0-E10-CS	DNA	pBluescriptII SK-
142	188-34-4-0-E5-CS	DNA	pBluescriptII SK-
143	188-9-3-0-A5-CS	DNA	pBluescriptII SK-
144	105-021-3-0-C3-CS	DNA	pBluescriptII SK-
145	105-037-4-0-H12-CS	DNA	pBluescriptII SK-
146	105-073-2-0-A7-CS	DNA	pBluescriptII SK-
147	109-002-4-0-C6-CS	DNA	pBluescriptII SK-
148	109-003-1-0-G4-CS	DNA	pBluescriptII SK-
149	116-118-4-0-A8-CS	DNA	pBluescriptII SK-
150	145-52-2-0-D12-CS	DNA	pBluescriptII SK-
151	145-7-2-0-G5-CS	DNA	pBluescriptII SK-
152	145-7-3-0-D3-CS	DNA	pBluescriptII SK-

153	157-17-2-0-C1-CS	DNA	pBluescriptII SK-
154	160-101-3-0-H2-CS	DNA	pBluescriptII SK-
155	160-12-1-0-D10-CS	DNA	pBluescriptII SK-
156	160-28-4-0-C4-CS	DNA	pBluescriptII SK-
157	160-31-3-0-E4-CS	DNA	pBluescriptII SK-
158	160-40-1-0-H4-CS	DNA	pBluescriptII SK-
159	160-54-1-0-F7-CS	DNA	pBluescriptII SK-
160	160-88-3-0-A8-CS.cor	DNA	pBluescriptII SK-
161	160-88-3-0-A8-CS.fr	DNA	pBluescriptII SK-
162	160-99-4-0-E4-CS	DNA	pBluescriptII SK-
163	161-5-4-0-B6-CS	DNA	pBluescriptII SK-
164	174-17-1-0-D6-CS	DNA	pPT
165	174-32-4-0-F8-CS	DNA	pPT
166	174-38-4-0-D11-CS	DNA	pPT
167	174-8-2-0-C10-CS	DNA	pPT
168	179-14-2-0-F11-CS	DNA	pBluescriptII SK-
169	179-9-4-0-B8-CS	DNA	pBluescriptII SK-
170	181-10-1-0-C9-CS	DNA	pBluescriptII SK-
171	187-5-3-0-C7-CS	DNA	pBluescriptII SK-
172	188-26-4-0-F5-CS	DNA	pBluescriptII SK-
173	188-27-3-0-G1-CS	DNA	pBluescriptII SK-
174	188-29-2-0-H1-CS	DNA	pBluescriptII SK-
175	188-31-1-0-E6-CS	DNA	pBluescriptII SK-
176	188-45-1-0-D3-CS	DNA	pBluescriptII SK-
177	188-5-1-0-H6-CS	DNA	pBluescriptII SK-
178	188-9-1-0-C10-CS	DNA	pBluescriptII SK-
179	105-016-3-0-C5-CS	DNA	pBluescriptII SK-
180	105-026-4-0-D9-CS	DNA	pBluescriptII SK-
181	105-053-2-0-D9-CS	DNA	pBluescriptII SK-
182	105-069-3-0-A11-CS	DNA	pBluescriptII SK-
183	105-076-4-0-F6-CS	DNA	pBluescriptII SK-
184	105-135-2-0-F9-CS	DNA	pBluescriptII SK-
185	106-023-4-0-F6-CS	DNA	pBluescriptII SK-
186	110-001-3-0-C11-CS	DNA	pBluescriptII SK-
187	110-002-3-0-F9-CS	DNA	pBluescriptII SK-
188	114-019-3-0-D9-CS	DNA	pBluescriptII SK-
189	114-029-1-0-C6-CS	DNA	pBluescriptII SK-
190	114-032-4-0-B1-CS	DNA	pBluescriptII SK-
191	114-070-2-0-H4-CS	DNA	pBluescriptII SK-
192	116-016-3-0-F11-CS	DNA	pBluescriptII SK-
193	116-022-4-0-G2-CS	DNA	pBluescriptII SK-
194	116-052-2-0-H8-CS	DNA	pBluescriptII SK-
195	116-053-4-0-B4-CS	DNA	pBluescriptII SK-
196	116-094-3-0-H2-CS	DNA	pBluescriptII SK-
197	116-112-4-0-C7-CS	DNA	pBluescriptII SK-
198	116-123-3-0-F12-CS	DNA	pBluescriptII SK-
199	123-008-1-0-C5-CS	DNA	pBluescriptII SK-
200	145-53-2-0-H8-CS	DNA	pBluescriptII SK-
201	145-57-2-0-C9-CS.cor	DNA	pBluescriptII SK-
202	145-57-2-0-C9-CS.fr	DNA	pBluescriptII SK-
203	145-7-3-0-B12-CS	DNA	pBluescriptII SK-
204	157-12-2-0-D1-CS	DNA	pBluescriptII SK-

205	157-16-2-0-D5-CS	DNA	pBluescriptII SK-
206	157-18-2-0-A7-CS	DNA	pBluescriptII SK-
207	160-103-1-0-B10-CS	DNA	pBluescriptII SK-
208	160-104-4-0-F3-CS	DNA	pBluescriptII SK-
209	160-22-2-0-D10-CS	DNA	pBluescriptII SK-
210	160-24-3-0-F12-CS	DNA	pBluescriptII SK-
211	160-3-2-0-H3-CS	DNA	pBluescriptII SK-
212	160-58-2-0-A2-CS	DNA	pBluescriptII SK-
213	160-73-1-0-B4-CS	DNA	pBluescriptII SK-
214	160-75-4-0-E6-CS	DNA	pBluescriptII SK-
215	160-97-3-0-E9-CS	DNA	pBluescriptII SK-
216	174-1-4-0-E9-CS	DNA	pPT
217	174-12-4-0-C2-CS	DNA	pPT
218	180-19-4-0-H2-CS	DNA	pBluescriptII SK-
219	181-10-4-0-G12-CS	DNA	pBluescriptII SK-
220	181-3-2-0-F6-CS	DNA	pBluescriptII SK-
221	181-4-4-0-A12-CS	DNA	pBluescriptII SK-
222	181-9-2-0-F12-CS.cor	DNA	pBluescriptII SK-
223	181-9-2-0-F12-CS.fr	DNA	pBluescriptII SK-
224	184-13-3-0-E11-CS	DNA	pBluescriptII SK-
225	184-4-2-0-D3-CS	DNA	pBluescriptII SK-
226	184-7-1-0-E7-CS	DNA	pBluescriptII SK-
227	184-8-4-0-G9-CS	DNA	pBluescriptII SK-
228	187-10-3-0-G9-CS	DNA	pBluescriptII SK-
229	187-32-0-0-m20-CS	DNA	pBluescriptII SK-
230	187-32-0-0-n21-CS.cor	DNA	pBluescriptII SK-
231	187-32-0-0-n21-CS.fr	DNA	pBluescriptII SK-
232	187-4-2-0-E6-CS	DNA	pBluescriptII SK-
233	187-40-0-0-i15-CS	DNA	pBluescriptII SK-
234	187-47-0-0-g24-CS	DNA	pBluescriptII SK-
235	187-9-3-0-A2-CS	DNA	pBluescriptII SK-
236	188-26-4-0-H1-CS	DNA	pBluescriptII SK-
237	188-35-3-0-G9-CS	DNA	pBluescriptII SK-
238	188-38-4-0-D8-CS	DNA	pBluescriptII SK-
239	188-41-1-0-E6-CS	DNA	pBluescriptII SK-
240	188-42-2-0-F3-CS.cor	DNA	pBluescriptII SK-
241	188-42-2-0-F3-CS.fr	DNA	pBluescriptII SK-
242	119-003-4-0-C2-CS	PRT	pBluescriptII SK-
243	105-016-1-0-D3-CS	PRT	pBluescriptII SK-
244	105-016-3-0-G10-CS	PRT	pBluescriptII SK-
245	105-026-1-0-A5-CS	PRT	pBluescriptII SK-
246	105-031-1-0-A11-CS	PRT	pBluescriptII SK-
247	105-031-2-0-D3-CS	PRT	pBluescriptII SK-
248	105-035-2-0-C6-CS	PRT	pBluescriptII SK-
249	105-037-2-0-H11-CS	PRT	pBluescriptII SK-
250	105-053-4-0-E8-CS	PRT	pBluescriptII SK-
251	105-074-3-0-H10-CS	PRT	pBluescriptII SK-
252	105-089-3-0-G10-CS	PRT	pBluescriptII SK-
253	105-095-2-0-G11-CS	PRT	pBluescriptII SK-
254	106-006-1-0-E3-CS	PRT	pBluescriptII SK-
255	106-037-1-0-E9-CS.cor	PRT	pBluescriptII SK-
256	106-037-1-0-E9-CS.fr	PRT	pBluescriptII SK-

257	106-043-4-0-H3-CS	PRT	pBluescriptII SK-
258	110-007-1-0-C7-CS	PRT	pBluescriptII SK-
259	114-016-1-0-H8-CS	PRT	pBluescriptII SK-
260	116-004-3-0-A6-CS	PRT	pBluescriptII SK-
261	116-054-3-0-E6-CS	PRT	pBluescriptII SK-
262	116-055-1-0-A3-CS	PRT	pBluescriptII SK-
263	116-055-2-0-F7-CS	PRT	pBluescriptII SK-
264	116-088-4-0-A9-CS	PRT	pBluescriptII SK-
265	116-091-1-0-D9-CS	PRT	pBluescriptII SK-
266	116-110-2-0-F4-CS	PRT	pBluescriptII SK-
267	116-111-1-0-H9-CS	PRT	pBluescriptII SK-
268	116-111-4-0-B3-CS	PRT	pBluescriptII SK-
269	116-115-2-0-F8-CS	PRT	pBluescriptII SK-
270	116-119-3-0-H5-CS	PRT	pBluescriptII SK-
271	117-001-5-0-G3-CS	PRT	pBluescriptII SK-
272	145-25-3-0-B4-CS.cor	PRT	pBluescriptII SK-
273	145-25-3-0-B4-CS.fr	PRT	pBluescriptII SK-
274	145-56-3-0-D5-CS	PRT	pBluescriptII SK-
275	145-59-2-0-A7-CS	PRT	pBluescriptII SK-
276	157-15-4-0-B11-CS	PRT	pBluescriptII SK-
277	160-103-1-0-F11-CS	PRT	pBluescriptII SK-
278	160-37-2-0-H7-CS	PRT	pBluescriptII SK-
279	160-58-3-0-H3-CS	PRT	pBluescriptII SK-
280	160-75-4-0-A9-CS	PRT	pBluescriptII SK-
281	174-10-2-0-F8-CS	PRT	pPT
282	174-33-3-0-F6-CS	PRT	pPT
283	174-38-1-0-B6-CS	PRT	pPT
284	174-38-3-0-C9-CS	PRT	pPT
285	174-39-2-0-A3-CS	PRT	pPT
286	174-41-1-0-A6-CS	PRT	pPT
287	174-5-3-0-H7-CS	PRT	pPT
288	174-7-4-0-H1-CS	PRT	pPT
289	175-1-3-0-E5-CS.cor	PRT	pPT
290	175-1-3-0-E5-CS.fr	PRT	pPT
291	180-19-4-0-F4-CS	PRT	pBluescriptII SK-
292	181-10-1-0-D10-CS	PRT	pBluescriptII SK-
293	181-16-1-0-G7-CS	PRT	pBluescriptII SK-
294	181-16-2-0-A7-CS	PRT	pBluescriptII SK-
295	181-20-3-0-B5-CS	PRT	pBluescriptII SK-
296	181-3-3-0-B8-CS	PRT	pBluescriptII SK-
297	181-3-3-0-C9-CS	PRT	pBluescriptII SK-
298	182-1-2-0-D12-CS	PRT	pBluescriptII SK-
299	184-1-4-0-C11-CS	PRT	pBluescriptII SK-
300	184-4-1-0-A11-CS	PRT	pBluescriptII SK-
301	187-12-4-0-A8-CS	PRT	pBluescriptII SK-
302	187-2-2-0-A3-CS	PRT	pBluescriptII SK-
303	187-31-0-0-f12-CS	PRT	pBluescriptII SK-
304	187-34-0-0-l12-CS	PRT	pBluescriptII SK-
305	187-37-0-0-c10-CS	PRT	pBluescriptII SK-
306	187-38-0-0-l10-CS	PRT	pBluescriptII SK-
307	187-39-0-0-k12-CS	PRT	pBluescriptII SK-
308	187-41-0-0-i21-CS	PRT	pBluescriptII SK-

309	188-11-1-0-B3-CS	PRT	pBluescriptII SK-
310	188-18-4-0-A9-CS	PRT	pBluescriptII SK-
311	188-28-4-0-B12-CS.cor	PRT	pBluescriptII SK-
312	188-28-4-0-B12-CS.fr	PRT	pBluescriptII SK-
313	188-28-4-0-D4-CS	PRT	pBluescriptII SK-
314	188-41-1-0-B8-CS.cor	PRT	pBluescriptII SK-
315	188-41-1-0-B8-CS.fr	PRT	pBluescriptII SK-
316	188-45-1-0-D9-CS	PRT	pBluescriptII SK-
317	188-9-2-0-E1-CS	PRT	pBluescriptII SK-
318	105-079-3-0-A11-CS	PRT	pBluescriptII SK-
319	105-092-1-0-H7-CS	PRT	pBluescriptII SK-
320	105-141-4-0-H9-CS	PRT	pBluescriptII SK-
321	109-013-1-0-B9-CS	PRT	pBluescriptII SK-
322	110-008-4-0-D9-CS	PRT	pBluescriptII SK-
323	114-001-3-0-A2-CS	PRT	pBluescriptII SK-
324	114-028-2-0-C1-CS	PRT	pBluescriptII SK-
325	114-032-1-0-H10-CS	PRT	pBluescriptII SK-
326	114-043-2-0-A10-CS	PRT	pBluescriptII SK-
327	114-044-1-0-C5-CS	PRT	pBluescriptII SK-
328	116-003-3-0-D10-CS	PRT	pBluescriptII SK-
329	116-003-3-0-G12-CS	PRT	pBluescriptII SK-
330	116-011-2-0-F11-CS	PRT	pBluescriptII SK-
331	116-033-3-0-E4-CS	PRT	pBluescriptII SK-
332	116-041-4-0-B6-CS	PRT	pBluescriptII SK-
333	116-044-2-0-C4-CS	PRT	pBluescriptII SK-
334	116-075-1-0-E6-CS	PRT	pBluescriptII SK-
335	116-094-4-0-G5-CS	PRT	pBluescriptII SK-
336	117-005-3-0-F2-CS	PRT	pBluescriptII SK-
337	121-007-3-0-D9-CS	PRT	pBluescriptII SK-
338	145-91-3-0-D10-CS	PRT	pBluescriptII SK-
339	157-17-1-0-F4-CS	PRT	pBluescriptII SK-
340	160-11-3-0-G8-CS	PRT	pBluescriptII SK-
341	160-24-1-0-F12-CS	PRT	pBluescriptII SK-
342	160-24-2-0-E9-CS	PRT	pBluescriptII SK-
343	160-25-4-0-D2-CS	PRT	pBluescriptII SK-
344	160-31-3-0-A11-CS	PRT	pBluescriptII SK-
345	160-32-1-0-F6-CS	PRT	pBluescriptII SK-
346	160-37-1-0-A3-CS	PRT	pBluescriptII SK-
347	160-40-3-0-E9-CS	PRT	pBluescriptII SK-
348	160-58-3-0-E4-CS	PRT	pBluescriptII SK-
349	160-85-3-0-D4-CS	PRT	pBluescriptII SK-
350	160-95-3-0-A11-CS	PRT	pBluescriptII SK-
351	162-10-4-0-F9-CS.cor	PRT	pBluescriptII SK-
352	162-10-4-0-F9-CS.fr	PRT	pBluescriptII SK-
353	174-13-2-0-E4-CS	PRT	pPT
354	174-46-2-0-B11-CS	PRT	pPT
355	179-8-2-0-A6-CS	PRT	pBluescriptII SK-
356	180-22-3-0-B6-CS	PRT	pBluescriptII SK-
357	181-13-1-0-F7-CS	PRT	pBluescriptII SK-
358	181-15-4-0-F7-CS	PRT	pBluescriptII SK-
359	181-20-1-0-G7-CS	PRT	pBluescriptII SK-
360	184-15-3-0-D1-CS	PRT	pBluescriptII SK-

361	187-12-2-0-G11-CS	PRT	pBluescriptII SK-
362	187-2-2-0-A12-CS	PRT	pBluescriptII SK-
363	187-30-0-0-k23-CS	PRT	pBluescriptII SK-
364	187-36-0-0-e19-CS	PRT	pBluescriptII SK-
365	187-38-0-0-d22-CS	PRT	pBluescriptII SK-
366	187-39-0-0-b9-CS	PRT	pBluescriptII SK-
367	187-39-0-0-g6-CS	PRT	pBluescriptII SK-
368	187-45-0-0-l18-CS	PRT	pBluescriptII SK-
369	187-45-0-0-m21-CS	PRT	pBluescriptII SK-
370	187-45-0-0-n8-CS	PRT	pBluescriptII SK-
371	187-46-0-0-f23-CS	PRT	pBluescriptII SK-
372	187-5-1-0-A12-CS	PRT	pBluescriptII SK-
373	187-5-1-0-F6-CS	PRT	pBluescriptII SK-
374	187-5-2-0-B2-CS	PRT	pBluescriptII SK-
375	187-5-3-0-D5-CS	PRT	pBluescriptII SK-
376	187-51-0-0-f9-CS	PRT	pBluescriptII SK-
377	187-6-1-0-B9-CS	PRT	pBluescriptII SK-
378	187-6-4-0-C10-CS	PRT	pBluescriptII SK-
379	188-19-2-0-C8-CS	PRT	pBluescriptII SK-
380	188-22-4-0-G6-CS	PRT	pBluescriptII SK-
381	188-28-4-0-D11-CS	PRT	pBluescriptII SK-
382	188-29-1-0-E10-CS	PRT	pBluescriptII SK-
383	188-34-4-0-E5-CS	PRT	pBluescriptII SK-
384	188-9-3-0-A5-CS	PRT	pBluescriptII SK-
385	105-021-3-0-C3-CS	PRT	pBluescriptII SK-
386	105-037-4-0-H12-CS	PRT	pBluescriptII SK-
387	105-073-2-0-A7-CS	PRT	pBluescriptII SK-
388	109-002-4-0-C6-CS	PRT	pBluescriptII SK-
389	109-003-1-0-G4-CS	PRT	pBluescriptII SK-
390	116-118-4-0-A8-CS	PRT	pBluescriptII SK-
391	145-52-2-0-D12-CS	PRT	pBluescriptII SK-
392	145-7-2-0-G5-CS	PRT	pBluescriptII SK-
393	145-7-3-0-D3-CS	PRT	pBluescriptII SK-
394	157-17-2-0-C1-CS	PRT	pBluescriptII SK-
395	160-101-3-0-H2-CS	PRT	pBluescriptII SK-
396	160-12-1-0-D10-CS	PRT	pBluescriptII SK-
397	160-28-4-0-C4-CS	PRT	pBluescriptII SK-
398	160-31-3-0-E4-CS	PRT	pBluescriptII SK-
399	160-40-1-0-H4-CS	PRT	pBluescriptII SK-
400	160-54-1-0-F7-CS	PRT	pBluescriptII SK-
401	160-88-3-0-A8-CS.cor	PRT	pBluescriptII SK-
402	160-88-3-0-A8-CS.fr	PRT	pBluescriptII SK-
403	160-99-4-0-E4-CS	PRT	pBluescriptII SK-
404	161-5-4-0-B6-CS	PRT	pBluescriptII SK-
405	174-17-1-0-D6-CS	PRT	pPT
406	174-32-4-0-F8-CS	PRT	pPT
407	174-38-4-0-D11-CS	PRT	pPT
408	174-8-2-0-C10-CS	PRT	pPT
409	179-14-2-0-F11-CS	PRT	pBluescriptII SK-
410	179-9-4-0-B8-CS	PRT	pBluescriptII SK-
411	181-10-1-0-C9-CS	PRT	pBluescriptII SK-
412	187-5-3-0-C7-CS	PRT	pBluescriptII SK-

413	188-26-4-0-F5-CS	PRT	pBluescriptII SK-
414	188-27-3-0-G1-CS	PRT	pBluescriptII SK-
415	188-29-2-0-H1-CS	PRT	pBluescriptII SK-
416	188-31-1-0-E6-CS	PRT	pBluescriptII SK-
417	188-45-1-0-D3-CS	PRT	pBluescriptII SK-
418	188-5-1-0-H6-CS	PRT	pBluescriptII SK-
419	188-9-1-0-C10-CS	PRT	pBluescriptII SK-
420	105-016-3-0-C5-CS	PRT	pBluescriptII SK-
421	105-026-4-0-D9-CS	PRT	pBluescriptII SK-
422	105-053-2-0-D9-CS	PRT	pBluescriptII SK-
423	105-069-3-0-A11-CS	PRT	pBluescriptII SK-
424	105-076-4-0-F6-CS	PRT	pBluescriptII SK-
425	105-135-2-0-F9-CS	PRT	pBluescriptII SK-
426	106-023-4-0-F6-CS	PRT	pBluescriptII SK-
427	110-001-3-0-C11-CS	PRT	pBluescriptII SK-
428	110-002-3-0-F9-CS	PRT	pBluescriptII SK-
429	114-019-3-0-D9-CS	PRT	pBluescriptII SK-
430	114-029-1-0-C6-CS	PRT	pBluescriptII SK-
431	114-032-4-0-B1-CS	PRT	pBluescriptII SK-
432	114-070-2-0-H4-CS	PRT	pBluescriptII SK-
433	116-016-3-0-F11-CS	PRT	pBluescriptII SK-
434	116-022-4-0-G2-CS	PRT	pBluescriptII SK-
435	116-052-2-0-H8-CS	PRT	pBluescriptII SK-
436	116-053-4-0-B4-CS	PRT	pBluescriptII SK-
437	116-094-3-0-H2-CS	PRT	pBluescriptII SK-
438	116-112-4-0-C7-CS	PRT	pBluescriptII SK-
439	116-123-3-0-F12-CS	PRT	pBluescriptII SK-
440	123-008-1-0-C5-CS	PRT	pBluescriptII SK-
441	145-53-2-0-H8-CS	PRT	pBluescriptII SK-
442	145-57-2-0-C9-CS.cor	PRT	pBluescriptII SK-
443	145-57-2-0-C9-CS.fr	PRT	pBluescriptII SK-
444	145-7-3-0-B12-CS	PRT	pBluescriptII SK-
445	157-12-2-0-D1-CS	PRT	pBluescriptII SK-
446	157-16-2-0-D5-CS	PRT	pBluescriptII SK-
447	157-18-2-0-A7-CS	PRT	pBluescriptII SK-
448	160-103-1-0-B10-CS	PRT	pBluescriptII SK-
449	160-104-4-0-F3-CS	PRT	pBluescriptII SK-
450	160-22-2-0-D10-CS	PRT	pBluescriptII SK-
451	160-24-3-0-F12-CS	PRT	pBluescriptII SK-
452	160-3-2-0-H3-CS	PRT	pBluescriptII SK-
453	160-58-2-0-A2-CS	PRT	pBluescriptII SK-
454	160-73-1-0-B4-CS	PRT	pBluescriptII SK-
455	160-75-4-0-E6-CS	PRT	pBluescriptII SK-
456	160-97-3-0-E9-CS	PRT	pBluescriptII SK-
457	174-1-4-0-E9-CS	PRT	pPT
458	174-12-4-0-C2-CS	PRT	pPT
459	180-19-4-0-H2-CS	PRT	pBluescriptII SK-
460	181-10-4-0-G12-CS	PRT	pBluescriptII SK-
461	181-3-2-0-F6-CS	PRT	pBluescriptII SK-
462	181-4-4-0-A12-CS	PRT	pBluescriptII SK-
463	181-9-2-0-F12-CS.cor	PRT	pBluescriptII SK-
464	181-9-2-0-F12-CS.fr	PRT	pBluescriptII SK-

465	184-13-3-0-E11-CS	PRT	pBluescriptII SK-
466	184-4-2-0-D3-CS	PRT	pBluescriptII SK-
467	184-7-1-0-E7-CS	PRT	pBluescriptII SK-
468	184-8-4-0-G9-CS	PRT	pBluescriptII SK-
469	187-10-3-0-G9-CS	PRT	pBluescriptII SK-
470	187-32-0-0-m20-CS	PRT	pBluescriptII SK-
471	187-32-0-0-n21-CS.cor	PRT	pBluescriptII SK-
472	187-32-0-0-n21-CS.fr	PRT	pBluescriptII SK-
473	187-4-2-0-E6-CS	PRT	pBluescriptII SK-
474	187-40-0-0-i15-CS	PRT	pBluescriptII SK-
475	187-47-0-0-g24-CS	PRT	pBluescriptII SK-
476	187-9-3-0-A2-CS	PRT	pBluescriptII SK-
477	188-26-4-0-H1-CS	PRT	pBluescriptII SK-
478	188-35-3-0-G9-CS	PRT	pBluescriptII SK-
479	188-38-4-0-D8-CS	PRT	pBluescriptII SK-
480	188-41-1-0-E6-CS	PRT	pBluescriptII SK-
481	188-42-2-0-F3-CS.cor	PRT	pBluescriptII SK-
482	188-42-2-0-F3-CS.fr	PRT	pBluescriptII SK-

Table II

Seq Id N	Full coding sequence	Signal sequence	Coding sequence for mature protein	Polyadenylation signal	Polyadenylation site
1	[169-1692]	[169-249]	[250-1692]	[2126-2131]	[2152-2201]
2	[148-1140]	[148-240]	[241-1140]	[1592-1597]	[1615-1631]
3	[85-906]	[85-135]	[136-906]	[1159-1164]	[1184-1245]
4	[31-1248]	[31-135]	[136-1248]	None detected	[1607-1623]
5	[72-143]	[72-119]	[120-143]	[1416-1421]	[1438-1454]
6	[111-1154]	[111-197]	[198-1154]	[1602-1607]	[1623-1639]
7	[66-1256]	[66-173]	[174-1256]	None detected	[1752-1768]
8	[190-1398]	[190-252]	[253-1398]	[1470-1475]	[1494-1510]
9	[78-410]	[78-155]	[156-410]	None detected	[866-882]
10	[84-299]	[84-134]	[135-299]	[1814-1819]	[1833-1849]
11	[55-468]	[55-99]	[100-468]	[531-536]	[549-565]
12	[152-475]	[152-244]	[245-475]	[1623-1628]	[1647-1663]
13	[112-552]	[112-183]	[184-552]	[706-711]	[729-744]
14	[101-1243]	[101-199]	[200-1243]	[1720-1725]	[1745-1759]
15	[101-517]	[101-199]	[200-517]	[1716-1721]	[1741-1755]
16	[59-853]	[59-100]	[101-853]	[894-899]	[922-936]
17	[73-672]	[73-132]	[133-672]	[689-694]	[711-747]
18	[94-1275]	[94-210]	[211-1275]	[1849-1854]	[1870-1884]
19	[42-515]	[42-92]	[93-515]	[649-654]	[677-691]
20	[271-969]	[271-366]	[367-969]	[1093-1098]	[1124-1138]
21	[76-276]	[76-135]	[136-276]	[436-441]	[455-468]
22	[6-287]	[6-80]	[81-287]	[684-689]	[706-720]
23	[171-692]	[171-227]	[228-692]	[691-696]	[713-727]
24	[137-454]	[137-187]	[188-454]	[440-445]	[456-470]
25	[238-609]	[238-291]	[292-609]	[948-953]	[973-987]
26	[80-862]	[80-127]	[128-862]	[875-880]	[894-908]
27	[83-310]	[83-157]	[158-310]	[725-730]	[748-762]
28	[310-906]	[310-357]	[358-906]	[1071-1076]	[1088-1102]
29	[24-287]	[24-131]	[132-287]	[405-410]	[422-436]
30	[132-1574]	[132-206]	[207-1574]	[1899-1904]	[1923-1938]
31	[117-545]	[117-245]	[246-545]	None detected	[1100-1116]
32	[117-362]	none detected	[117-362]	None detected	[1098-1114]
33	[144-1262]	[144-224]	[225-1262]	[2035-2040]	[2056-2072]
34	[35-316]	[35-109]	[110-316]	None detected	[393-409]
35	[177-767]	[177-236]	[237-767]	None detected	[822-836]
36	[208-1239]	[208-294]	[295-1239]	None detected	[1307-1323]
37	[60-1682]	[60-143]	[144-1682]	None detected	[1929-1945]
38	[198-998]	[198-269]	[270-998]	[1292-1297]	[1315-1330]
39	[505-1590]	[505-624]	[625-1590]	[2089-2094]	[2108-2124]
40	[84-326]	[84-146]	[147-326]	[1122-1127]	[1142-1159]
41	[56-1678]	[56-139]	[140-1678]	None detected	[1936-1953]

42	[119-1522]	[119-181]	[182-1522]	None detected	[1671-1688]
43	[334-1551]	[334-426]	[427-1551]	None detected	[1925-1942]
44	[72-986]	[72-149]	[150-986]	[1608-1613]	[1640-1657]
45	[157-1482]	[157-219]	[220-1482]	None detected	[1716-1733]
46	[195-1052]	[195-338]	[339-1052]	None detected	[1854-1871]
47	[217-1410]	[217-279]	[280-1410]	[1482-1487]	[1507-1536]
48	[103-492]	[103-162]	[163-492]	[794-799]	[815-832]
49	[234-491]	[234-293]	[294-491]	[793-798]	[814-831]
50	[180-800]	[180-248]	[249-800]	[880-885]	[901-917]
51	[140-472]	[140-211]	[212-472]	None detected	[605-621]
52	[68-484]	[68-112]	[113-484]	None detected	[657-673]
53	[38-517]	[38-118]	[119-517]	[861-866]	[885-897]
54	[92-634]	[92-139]	[140-634]	None detected	None detected
55	[27-767]	[27-80]	[81-767]	None detected	[1031-1047]
56	[4-399]	[4-126]	[127-399]	[891-896]	[909-923]
57	[127-879]	[127-198]	[199-879]	None detected	[1224-1240]
58	[156-566]	[156-221]	[222-566]	[870-875]	[888-902]
59	[35-1657]	[35-118]	[119-1657]	None detected	[1955-1969]
60	[77-937]	[77-127]	[128-937]	[1098-1103]	[1116-1132]
61	[9-503]	[9-113]	[114-503]	[594-599]	[615-631]
62	[21-464]	[21-95]	[96-464]	[650-655]	[692-722]
63	[178-1050]	[178-279]	[280-1050]	[1400-1405]	[1426-1442]
64	[32-274]	[32-178]	[179-274]	[756-761]	[779-795]
65	[222-920]	[222-311]	[312-920]	[1191-1196]	[1220-1236]
66	[101-355]	[101-160]	[161-355]	[772-777]	[788-881]
67	[173-487]	[173-301]	[302-487]	[486-491]	[508-524]
68	[210-1082]	[210-311]	[312-1082]	[1432-1437]	[1456-1472]
69	[172-1449]	[172-255]	[256-1449]	None detected	[1721-1737]
70	[30-1427]	[30-77]	[78-1427]	[1594-1599]	[1621-1637]
71	[30-1175]	[30-77]	[78-1175]	[1593-1598]	[1620-1636]
72	[66-839]	[66-173]	[174-839]	None detected	[1742-1758]
73	[64-903]	[64-162]	[163-903]	[1612-1617]	[1631-1647]
74	[64-585]	[64-162]	[163-585]	[1611-1616]	[1630-1646]
75	[274-753]	[274-324]	[325-753]	[1931-1936]	[1947-1963]
76	[191-1468]	[191-274]	[275-1468]	None detected	[1741-1757]
77	[48-950]	[48-107]	[108-950]	[1983-1988]	[2011-2027]
78	[156-512]	[156-206]	[207-512]	[1831-1836]	[1864-1880]
79	[67-351]	[67-183]	[184-351]	None detected	[568-584]
80	[259-831]	[259-375]	[376-831]	None detected	[1337-1351]
81	[111-377]	[111-233]	[234-377]	[689-694]	[706-720]
82	[223-432]	[223-336]	[337-432]	[986-991]	[1015-1029]
83	[769-1272]	[769-843]	[844-1272]	None detected	[1774-1788]
84	[30-527]	[30-74]	[75-527]	[738-743]	[756-805]
85	[39-506]	[39-83]	[84-506]	None detected	[800-814]
86	[115-429]	[115-210]	[211-429]	[565-570]	[584-598]
87	[332-574]	[332-412]	[413-574]	None detected	[630-699]

88	[133-417]	[133-213]	[214-417]	[876-881]	[891-905]
89	[113-364]	[113-172]	[173-364]	None detected	[500-514]
90	[9-380]	[9-104]	[105-380]	[483-488]	[504-518]
91	[155-340]	[155-292]	[293-340]	[728-733]	[754-808]
92	[185-634]	[185-253]	[254-634]	[704-709]	[723-737]
93	[53-646]	[53-91]	[92-646]	[694-699]	[714-728]
94	[247-510]	[247-318]	[319-510]	[544-549]	[568-582]
95	[143-592]	[143-277]	[278-592]	[1877-1882]	[1898-1913]
96	[33-458]	[33-89]	[90-458]	[637-642]	[654-670]
97	[1-336]	[1-81]	[82-336]	[900-905]	[923-939]
98	[174-443]	[174-269]	[270-443]	[629-634]	[647-661]
99	[282-521]	[282-386]	[387-521]	[600-605]	[631-647]
100	[251-643]	[251-295]	[296-643]	None detected	[990-1006]
101	[179-475]	[179-295]	[296-475]	[995-1000]	[1015-1059]
102	[34-327]	[34-162]	[163-327]	[466-471]	[498-514]
103	[303-953]	[303-359]	[360-953]	[1124-1129]	[1142-1158]
104	[97-645]	[97-156]	[157-645]	[1524-1529]	[1547-1563]
105	[80-820]	[80-118]	[119-820]	[1587-1592]	[1606-1621]
106	[77-388]	[77-217]	[218-388]	[524-529]	[541-557]
107	[139-513]	[139-201]	[202-513]	[566-571]	[584-600]
108	[81-986]	[81-134]	[135-986]	[1092-1097]	[1113-1129]
109	[266-586]	[266-307]	[308-586]	[745-750]	[762-778]
110	[59-745]	[59-160]	[161-745]	None detected	[1285-1301]
111	[59-676]	[59-160]	[161-676]	None detected	[1284-1300]
112	[15-278]	[15-146]	[147-278]	[1580-1585]	[1600-1617]
113	[167-619]	[167-262]	[263-619]	[1598-1603]	[1617-1634]
114	[223-417]	[223-270]	[271-417]	[655-660]	[677-693]
115	[166-732]	[166-237]	[238-732]	[753-758]	[768-784]
116	[75-623]	[75-215]	[216-623]	[767-772]	[788-804]
117	[30-335]	[30-71]	[72-335]	[450-455]	[468-484]
118	[21-752]	[21-107]	[108-752]	None detected	[970-985]
119	[185-715]	[185-253]	[254-715]	[785-790]	[814-839]
120	[54-527]	[54-116]	[117-527]	[545-550]	[567-583]
121	[129-686]	[129-185]	[186-686]	[989-994]	[1008-1024]
122	[165-614]	[165-305]	[306-614]	[719-724]	[744-760]
123	[192-476]	[192-326]	[327-476]	[555-560]	[578-594]
124	[16-297]	[16-93]	[94-297]	None detected	[543-559]
125	[216-635]	[216-335]	[336-635]	[717-722]	[728-744]
126	[164-280]	[164-268]	[269-280]	[789-794]	[809-824]
127	[68-301]	[68-190]	[191-301]	[485-490]	[510-526]
128	[179-427]	[179-298]	[299-427]	[579-584]	[602-618]
129	[22-297]	[22-66]	[67-297]	[742-747]	[760-776]
130	[9-845]	[9-134]	[135-845]	[964-969]	[983-998]
131	[27-578]	[27-119]	[120-578]	[742-747]	[763-779]
132	[408-710]	[408-533]	[534-710]	[985-990]	[1009-1025]
133	[247-501]	[247-306]	[307-501]	None detected	[592-607]

134	[333-602]	[333-416]	[417-602]	None detected	[761-774]
135	[110-376]	[110-208]	[209-376]	[582-587]	[601-611]
136	[22-417]	[22-66]	[67-417]	[888-893]	[909-925]
137	[62-367]	[62-103]	[104-367]	[638-643]	[658-674]
138	[107-1618]	[107-178]	[179-1618]	[1688-1693]	[1709-1725]
139	[16-471]	[16-93]	[94-471]	None detected	[1458-1474]
140	[222-374]	[222-299]	[300-374]	None detected	[637-653]
141	[59-274]	[59-127]	[128-274]	[1452-1457]	[1474-1490]
142	[158-442]	[158-301]	[302-442]	[621-626]	[645-661]
143	[5-454]	[5-64]	[65-454]	[1745-1750]	[1773-1789]
144	[241-1302]	none detected	[241-1302]	[1968-1973]	[1990-2006]
145	[15-635]	none detected	[15-635]	[1057-1062]	[1080-1096]
146	[109-738]	none detected	[109-738]	[1633-1638]	[1650-1666]
147	[21-1145]	none detected	[21-1145]	[1648-1653]	[1666-1687]
148	[70-1596]	none detected	[70-1596]	[1712-1717]	[1733-1747]
149	[129-362]	none detected	[129-362]	[597-602]	[626-658]
150	[109-594]	none detected	[109-594]	[1999-2004]	[2029-2045]
151	[150-587]	none detected	[150-587]	None detected	[772-788]
152	[173-847]	none detected	[173-847]	[1894-1899]	[1915-1931]
153	[100-441]	none detected	[100-441]	[479-484]	[500-514]
154	[32-1132]	none detected	[32-1132]	None detected	[1167-1183]
155	[160-996]	none detected	[160-996]	[1504-1509]	[1529-1545]
156	[11-529]	none detected	[11-529]	[1042-1047]	[1053-1068]
157	[135-749]	none detected	[135-749]	[1055-1060]	[1081-1097]
158	[98-637]	none detected	[98-637]	[862-867]	[878-894]
159	[221-670]	none detected	[221-670]	[669-674]	[688-703]
160	[165-674]	none detected	[165-674]	[808-813]	[833-849]
161	[165-671]	none detected	[165-671]	[805-810]	[830-846]
162	[28-1128]	none detected	[28-1128]	[1121-1126]	[1159-1176]
163	[135-194]	none detected	[135-194]	[1050-1055]	[1068-1084]
164	[173-847]	none detected	[173-847]	[1757-1762]	[1776-1793]
165	[8-1141]	none detected	[8-1141]	None detected	[1832-1849]
166	[136-264]	none detected	[136-264]	[1720-1725]	[1731-1748]
167	[14-1048]	none detected	[14-1048]	[1234-1239]	[1258-1275]
168	[70-777]	none detected	[70-777]	[987-992]	[1007-1023]
169	[38-400]	none detected	[38-400]	[1043-1048]	[1069-1085]
170	[63-572]	none detected	[63-572]	[750-755]	[767-776]
171	[160-867]	none detected	[160-867]	[1178-1183]	[1203-1219]
172	[68-640]	none detected	[68-640]	None detected	[1471-1487]
173	[132-1298]	none detected	[132-1298]	[1873-1878]	[1899-1915]
174	[259-1701]	none detected	[259-1701]	None detected	[1974-1990]
175	[213-1274]	none detected	[213-1274]	[1940-1945]	[1955-1971]
176	[68-127]	none detected	[68-127]	None detected	[1597-1613]
177	[65-1024]	none detected	[65-1024]	[1291-1296]	[1315-1361]
178	[109-585]	none detected	[109-585]	[1059-1064]	[1082-1113]
179	[29-577]	none detected	[29-577]	[1917-1922]	[1944-1960]

180	[23-451]	none detected	[23-451]	[1405-1410]	[1427-1443]
181	[232-450]	none detected	[232-450]	None detected	[589-605]
182	[758-1183]	none detected	[758-1183]	None detected	[1708-1724]
183	[486-932]	none detected	[486-932]	None detected	[1670-1686]
184	[80-304]	none detected	[80-304]	None detected	[452-463]
185	[188-691]	none detected	[188-691]	[707-712]	[727-773]
186	[94-573]	none detected	[94-573]	None detected	[739-753]
187	[181-462]	none detected	[181-462]	None detected	[740-754]
188	[6-290]	none detected	[6-290]	None detected	[971-998]
189	[115-411]	none detected	[115-411]	[573-578]	[591-605]
190	[3-368]	none detected	[3-368]	[481-486]	[511-526]
191	[174-527]	none detected	[174-527]	[878-883]	[896-910]
192	[57-203]	none detected	[57-203]	[579-584]	[599-668]
193	[68-334]	none detected	[68-334]	[562-567]	[583-637]
194	[183-443]	none detected	[183-443]	[670-675]	[692-706]
195	[94-228]	none detected	[94-228]	None detected	[656-670]
196	[133-327]	none detected	[133-327]	[465-470]	[496-510]
197	[22-357]	none detected	[22-357]	None detected	[486-500]
198	[4-333]	none detected	[4-333]	[633-638]	[653-667]
199	[1-363]	none detected	[1-363]	[474-479]	[498-514]
200	[41-337]	none detected	[41-337]	None detected	[401-462]
201	[1-551]	none detected	[1-551]	None detected	[535-551]
202	[34-315]	none detected	[34-315]	None detected	[534-550]
203	[1-315]	none detected	[1-315]	[371-376]	[392-408]
204	[94-582]	none detected	[94-582]	None detected	[651-665]
205	[540-923]	none detected	[540-923]	None detected	[994-1008]
206	[77-364]	none detected	[77-364]	[367-372]	[391-455]
207	[65-544]	none detected	[65-544]	[710-715]	[733-749]
208	[117-467]	none detected	[117-467]	[557-562]	[578-594]
209	[893-1897]	none detected	[893-1897]	[2066-2071]	[2082-2098]
210	[85-342]	none detected	[85-342]	None detected	[412-428]
211	[155-433]	none detected	[155-433]	[713-718]	[735-769]
212	[63-386]	none detected	[63-386]	[878-883]	[898-914]
213	[460-1290]	none detected	[460-1290]	[1449-1454]	[1473-1489]
214	[21-539]	none detected	[21-539]	[741-746]	[760-776]
215	[34-1143]	none detected	[34-1143]	[1375-1380]	[1397-1412]
216	[6-1184]	none detected	[6-1184]	[1735-1740]	[1744-1773]
217	[29-376]	none detected	[29-376]	None detected	[1184-1251]
218	[78-566]	none detected	[78-566]	[858-863]	[878-894]
219	[16-705]	none detected	[16-705]	[868-873]	[894-910]
220	[103-405]	none detected	[103-405]	[482-487]	[503-519]
221	[72-350]	none detected	[72-350]	[593-598]	[616-632]
222	[38-436]	none detected	[38-436]	None detected	[636-652]
223	[38-322]	none detected	[38-322]	None detected	[634-650]
224	[202-480]	none detected	[202-480]	[472-477]	[488-502]
225	[171-1670]	none detected	[171-1670]	[1706-1711]	[1725-1739]

226	[199-618]	none detected	[199-618]	[626-631]	[643-657]
227	[182-481]	none detected	[182-481]	None detected	[874-888]
228	[161-517]	none detected	[161-517]	None detected	[701-716]
229	[86-505]	none detected	[86-505]	[618-623]	[638-654]
230	[56-382]	none detected	[56-382]	[598-603]	[619-635]
231	[56-355]	none detected	[56-355]	[597-602]	[618-634]
232	[76-498]	none detected	[76-498]	[546-551]	[567-583]
233	[199-600]	none detected	[199-600]	[705-710]	[737-753]
234	[211-612]	none detected	[211-612]	[717-722]	[746-762]
235	[5-259]	none detected	[5-259]	[502-507]	[521-537]
236	[23-370]	none detected	[23-370]	[956-961]	[978-994]
237	[41-352]	none detected	[41-352]	None detected	[646-662]
238	[3-1319]	none detected	[3-1319]	[1791-1796]	[1813-1829]
239	[421-768]	none detected	[421-768]	[1045-1050]	[1067-1083]
240	[78-590]	none detected	[78-590]	None detected	[1815-1831]
241	[78-608]	none detected	[78-608]	None detected	[1814-1830]

Table III

List of variants
92;119
14;15
110;111
69;174;76
2;12
172;176;177
150;152;164;166
154;162
77;143
34;62
230;231
63;68
8;47
48;49;66
7;72
160;161
144;175
17;21
31;32
5;6
3;10
96;121
37;41;59
70;71
19;24
186;195;204
73;74
240;241
221;235
222;223
42;45
157;163
190;229
117;137
122;233;234
201;202
80;139

Table IV

Seq Id No	Preferentially excluded fragments
1	192..235;2099..2201
2	174..225;1605..1631
3	1111..1245
4	1590..1598;1607..1623
5	1385..1453
6	1571..1639
7	1732..1768
8	1494..1510
9	570..882
10	1176..1218;1710..1742;1833..1849
11	219..253;455..565
12	178..229;1636..1663
13	729..744
14	790..827;1735..1759
15	788..825;1731..1755
16	922..936
17	668..747
18	1870..1884
19	677..691
20	1124..1138
21	450..468
22	393..411;706..720
23	713..727
24	456..470
25	876..928;973..987
26	894..908
27	748..762
28	1088..1102
29	422..436
30	1879..1918;1923..1938
31	774..1116
32	772..1114
33	2056..2072
34	393..409
35	784..836
36	544..551;1307..1323
37	1867..1874;1929..1945
38	1315..1330
39	2108..2124
40	413..421;1116..1159
41	1863..1870;1936..1953
42	1623..1688

43	1895..1942
44	1640..1657
45	1661..1733
46	1555..1871
47	1507..1523
48	541..832
49	540..831
50	901..917
51	2..10;605..621
52	585..673
53	885..897
54	4..13;761..1101
55	1031..1047
56	873..905;907..923
57	1224..1240
58	861..902
59	1842..1849;1955..1969
60	1116..1132
61	15..46;615..631
62	651..722
63	1426..1442
64	739..795
65	1220..1236
66	520..881
67	413..524
68	1444..1472
69	1721..1737
70	1621..1637
71	1620..1636
72	777..784;1742..1758
73	1631..1647
74	1630..1646
75	1947..1963
76	1741..1757
77	1561..1913;2011..2027
78	727..819;880..894;901..1280;1841..1880
79	418..584
80	331..353;844..1214;1337..1351
81	706..720
82	639..713;1008..1029
83	1454..1788
84	712..805
85	800..814
86	584..598
87	122..308;593..699
88	855..905

89	500..514
90	81..101;198..205;504..518
91	650..808
92	128..201;723..737
93	714..728
94	568..582
95	1761..1773;1898..1913
96	654..670
97	883..938
98	616..661
99	631..647
100	853..1006
101	537..544;949..1059
102	498..514
103	1142..1158
104	1524..1563
105	1230..1259;1606..1621
106	505..557
107	584..600
108	378..385;1113..1129
109	729..778
110	992..1301
111	991..1300
112	1131..1139;1569..1617
113	1526..1634
114	457..509;677..693
115	768..784
116	360..670;788..804
117	435..484
118	433..452;764..985
119	128..201;801..839
120	554..564;567..583
121	872..908;1008..1024
122	744..760
123	578..594
124	94..102;248..559
125	728..744
126	809..824
127	510..526
128	602..618
129	472..553;569..776
130	983..998
131	396..468;763..779
132	478..532;1009..1025
133	592..607
134	761..774

135	556..563;601..611
136	887..919
137	658..674
138	1651..1725
139	49..71;988..1358;1458..1474
140	324..653
141	720..730;1449..1490
142	44..119;498..505;578..585;645..661
143	1322..1666;1773..1789
144	1828..1897;1919..1968;1990..2006
145	936..955;1060..1096
146	778..827;1650..1666
147	1170..1207;1647..1687
148	1733..1747
149	579..658
150	1432..1440;1728..1778;2004..2045
151	772..788
152	1496..1504;1792..1842;1915..1931
153	500..514
154	1167..1183
155	1529..1545
156	703..1068
157	873..881;1081..1097
158	878..894
159	688..703
160	833..849
161	830..846
162	1159..1176
163	869..876;1068..1084
164	1444..1463;1496..1504;1743..1793
165	1233..1319;1697..1849
166	1407..1426;1459..1467;1694..1748
167	1258..1275
168	84..129;1002..1023
169	436..472;596..604;673..689;732..954;995..1085
170	767..776
171	1203..1219
172	1411..1487
173	1861..1915
174	1974..1990
175	1800..1869;1891..1940;1955..1971
176	1597..1613
177	186..212;1277..1361
178	930..978;1002..1113
179	951..1000;1364..1533;1944..1960
180	1427..1443

181	107..181;276..311;449..605
182	1143..1450;1677..1724
183	1..251;648..655;1347..1686
184	447..463
185	150..159;623..773
186	340..476;739..753
187	740..754
188	307..315;668..998
189	118..125;529..536;591..605
190	492..526
191	872..910
192	525..668
193	91..135;461..637
194	392..458;551..671;692..706
195	656..670
196	283..379;458..466;496..510
197	1..96;483..500
198	625..667
199	474..513
200	370..462
201	535..551
202	534..550
203	374..408
204	651..665
205	994..1008
206	348..455
207	733..749
208	1..49;578..594
209	2082..2098
210	412..428
211	689..769
212	898..914
213	1266..1489
214	760..776
215	1304..1311;1383..1412
216	648..691;1711..1773
217	644..856;910..1251
218	878..894
219	894..910
220	503..519
221	616..632
222	636..652
223	634..650
224	50..57;488..502
225	534..577;1725..1739
226	643..657

227	1..84;874..888
228	701..716
229	638..654
230	263..573;619..635
231	263..573;619..635
232	567..583
233	737..753
234	746..762
235	499..537
236	905..912;944..994
237	348..662
239	829..1083
240	1508..1831
241	1507..1830

Table Va

Seq Id No	Preferentially excluded fragments	Preferentially included fragments
1	[1-540];[556-615];[2061-2096];[2098-2201]	[541-555];[616-2060];[2097-2097]
2	[1-511];[533-619];[621-690];[730-1132]	[512-532];[620-620];[691-729];[1133-1631]
3	[2-539];[1178-1245]	[1-1];[540-1177]
4	[1-250];[297-383];[386-514];[1025-1064]	[251-296];[384-385];[515-1024];[1065-1623]
5	[27-116];[118-391]	[1-26];[117-117];[392-1454]
6	[1-93];[96-168];[170-262];[264-461]	[94-95];[169-169];[263-263];[462-1639]
7	[1-95];[97-451]	[96-96];[452-1768]
8	[1-502];[1314-1491]	[503-1313];[1492-1510]
9	[1-864]	[865-882]
10	[1-428]	[429-1849]
11	[1-454];[482-514]	[455-481];[515-565]
12	[1-375];[379-511];[533-690];[730-783];[814-1164]	[376-378];[512-532];[691-729];[784-813];[1165-1663]
13	[2-337];[339-556]	[1-1];[338-338];[557-744]
14	[29-366];[368-507]	[1-28];[367-367];[508-1759]
15	[29-366];[368-524]	[1-28];[367-367];[525-1755]
16	[1-641]	[642-936]
17	[1-708];[711-747]	[709-710]
18	[1-639]	[640-1884]
19	[1-631]	[632-691]
20	[3-416];[418-490]	[1-2];[417-417];[491-1138]
21	[1-468]	None
22	[1-720]	None
23	[1-711]	[712-727]
24	[1-469]	[470-470]
25	[1-231];[234-488]	[232-233];[489-987]
26	[1-296];[300-642];[644-737]	[297-299];[643-643];[738-908]
27	[1-306];[308-762]	[307-307]
28	[1-446];[448-1102]	[447-447]
29	[1-436]	None
30	[7-334];[1420-1468];[1474-1614];[1616-1804];[1845-1919]	[1-6];[335-1419];[1469-1473];[1615-1615];[1805-1844];[1920-1938]
31	[1-342];[345-519];[823-893];[977-1016]	[343-344];[520-822];[894-976];[1017-1116]
32	[1-517];[821-891];[975-1014]	[518-820];[892-974];[1015-1114]
33	[36-352];[354-457];[728-832];[834-1096];[1253-1289];[1291-1350];[1352-1412];[1726-1873]	[1-35];[353-353];[458-727];[833-833];[1097-1252];[1290-1290];[1351-1351];[1413-1725];[1874-2072]
34	[1-409]	None
35	[14-105]	[1-13];[106-836]
36	[1-572];[1120-1271]	[573-1119];[1272-1323]
37	[20-98];[100-510];[1591-1681];[1683-1870]	[1-19];[99-99];[511-1590];[1682-1682];[1871-1945]

38	[1-547]	[548-1330]
39	[1-445]	[446-2124]
40	[1-473];[475-528]	[474-474];[529-1159]
41	[16-506];[1587-1866]	[1-15];[507-1586];[1867-1953]
42	[2-234];[244-451];[974-1226]	[1-1];[235-243];[452-973];[1227-1688]
43	[1-455];[1670-1925]	[456-1669];[1926-1942]
44	[1-579];[815-1031]	[580-814];[1032-1657]
45	[1-489];[1012-1264]	[490-1011];[1265-1733]
46	[1-400];[1184-1223];[1225-1705];[1740-1818]	[401-1183];[1224-1224];[1706-1739];[1819-1871]
47	[1-529];[1326-1505]	[530-1325];[1506-1523]
48	[1-131];[133-510];[560-589]	[132-132];[511-559];[590-832]
49	[1-130];[132-509];[559-588]	[131-131];[510-558];[589-831]
50	[1-650];[652-868];[873-913]	[651-651];[869-872];[914-917]
51	[1-504];[515-605]	[505-514];[606-621]
52	[1-535]	[536-673]
53	[2-563]	[1-1];[564-897]
54	[1-527];[802-870];[882-934];[966-1018];[1037-1080]	[528-801];[871-881];[935-965];[1019-1036];[1081-1101]
55	[1-326];[328-505]	[327-327];[506-1047]
56	[1-340]	[341-925]
57	[1-528]	[529-1240]
58	[1-108];[115-151];[154-340];[342-529]	[109-114];[152-153];[341-341];[530-902]
59	[4-485];[1566-1656];[1658-1845]	[1-3];[486-1565];[1657-1657];[1846-1969]
60	[1-283]	[284-1132]
61	[9-468]	[1-8];[469-631]
62	[1-525];[689-722]	[526-688]
63	[1-88];[90-192];[194-265];[296-409]	[89-89];[193-193];[266-295];[410-1442]
64	[1-517]	[518-795]
65	[1-406];[408-739]	[407-407];[740-1236]
66	[1-489];[849-881]	[490-848]
67	[1-505]	[506-524]
68	[1-325];[328-441];[444-504]	[326-327];[442-443];[505-1472]
69	[1-524];[636-715];[717-809];[811-885];[1567-1715]	[525-635];[716-716];[810-810];[886-1566];[1716-1737]
70	[12-487]	[1-11];[488-1637]
71	[12-487]	[1-11];[488-1636]
72	[1-451]	[452-1758]
73	[1-167];[242-464]	[168-241];[465-1647]
74	[1-167];[242-464]	[168-241];[465-1646]
75	[1-471]	[472-1963]
76	[1-358];[360-543];[655-734];[736-828];[830-904];[1586-1734]	[359-359];[544-654];[735-735];[829-829];[905-1585];[1735-1757]
77	[3-34];[36-474];[582-770];[1709-1746];[1748-1785];[1825-1899]	[1-2];[35-35];[475-581];[771-1708];[1747-1747];[1786-1824];[1900-2027]
78	[1-75];[77-319];[914-1052];[1063-1126];[1168-1203]	[76-76];[320-913];[1053-1062];[1127-1167];[1204-1880]

79	[1-425]	[426-584]
80	[1-752];[947-1017];[1084-1170]	[753-946];[1018-1083];[1171-1351]
81	[1-496];[498-720]	[497-497]
82	[1-324]	[325-1029]
83	[1-477];[1474-1529];[1537-1566];[1577-1616];[1622-1662];[1717-1753]	[478-1473];[1530-1536];[1567-1576];[1617-1621];[1663-1716];[1754-1788]
84	[1-496];[499-568];[752-805]	[497-498];[569-751]
85	[1-527]	[528-814]
86	[1-360]	[361-598]
87	[1-78];[80-583];[625-699]	[79-79];[584-624]
88	[1-889]	[890-905]
89	[1-513]	[514-514]
90	[1-122];[124-155];[157-435];[437-517]	[123-123];[156-156];[436-436];[518-518]
91	[1-133];[165-808]	[134-164]
92	[1-725]	[726-737]
93	[1-409]	[410-728]
94	[1-331]	[332-582]
95	[1-410]	[411-1913]
96	[1-501]	[502-670]
97	[1-141];[143-431]	[142-142];[432-939]
98	[1-193]	[194-661]
99	[1-629]	[630-647]
100	[1-520];[862-954];[976-1005]	[521-861];[955-975];[1006-1006]
101	[1-489];[581-961];[1010-1059]	[490-580];[962-1009]
102	[1-485]	[486-514]
103	[1-540]	[541-1158]
104	[1-556]	[557-1563]
105	[1-868];[870-1006]	[869-869];[1007-1621]
106	[1-491]	[492-557]
107	[1-573]	[574-600]
108	[1-457];[586-1110]	[458-585];[1111-1129]
109	[1-521];[655-778]	[522-654]
110	[1-416];[478-614];[616-990];[992-1065];[1068-1283]	[417-477];[615-615];[991-991];[1066-1067];[1284-1301]
111	[1-416];[478-614];[628-989];[991-1064];[1067-1282]	[417-477];[615-627];[990-990];[1065-1066];[1283-1300]
112	[2-429];[1161-1202];[1212-1388];[1392-1589]	[1-1];[430-1160];[1203-1211];[1389-1391];[1590-1617]
113	[1-487]	[488-1634]
114	[1-70];[86-496]	[71-85];[497-693]
115	[1-358];[360-558]	[359-359];[559-784]
116	[1-215];[218-495];[527-607]	[216-217];[496-526];[608-804]
117	[1-466]	[467-484]
118	[1-515];[906-963]	[516-905];[964-985]
119	[1-744];[746-816]	[745-745];[817-839]
120	[1-85];[87-521]	[86-86];[522-583]

121	[1-532]	[533-1024]
122	[1-318];[325-517];[567-660]	[319-324];[518-566];[661-760]
123	[1-498]	[499-594]
124	[1-427]	[428-559]
125	[1-642]	[643-744]
126	[1-341];[350-696]	[342-349];[697-824]
127	[1-482]	[483-526]
128	[1-338]	[339-618]
129	[1-191];[193-429];[450-678]	[192-192];[430-449];[679-776]
130	[19-463];[465-544]	[1-18];[464-464];[545-998]
131	[1-470]	[471-779]
132	[1-533]	[534-1025]
133	[1-498]	[499-607]
134	[1-168];[170-326];[328-471];[552-738]	[169-169];[327-327];[472-551];[739-774]
135	[1-346];[348-395];[440-473]	[347-347];[396-439];[474-611]
136	[1-324];[343-436]	[325-342];[437-925]
137	[1-186];[188-251];[255-517]	[187-187];[252-254];[518-674]
138	[1-488]	[489-1725]
139	[1-101];[103-190];[292-327];[1091-1161];[1228-1314]	[102-102];[191-291];[328-1090];[1162-1227];[1315-1474]
140	[1-465];[516-653]	[466-515]
141	[1-761];[763-857];[912-1326]	[762-762];[858-911];[1327-1490]
142	[1-476]	[477-661]
143	[1-531];[1471-1508];[1510-1547];[1587-1661]	[532-1470];[1509-1509];[1548-1586];[1662-1789]
144	[1-492];[503-536]	[493-502];[537-2006]
145	[1-570]	[571-1096]
146	[1-536];[621-703];[729-1075];[1198-1445]	[537-620];[704-728];[1076-1197];[1446-1666]
147	[1-555];[578-628]	[556-577];[629-1687]
148	[1-444];[1201-1474];[1480-1516]	[445-1200];[1475-1479];[1517-1747]
149	[1-613];[626-658]	[614-625]
150	[4-199];[201-419];[421-492]	[1-3];[200-200];[420-420];[493-2045]
151	[1-509]	[510-788]
152	[1-483];[485-578]	[484-484];[579-1931]
153	[1-497]	[498-514]
154	[5-509];[579-763];[765-1162]	[1-4];[510-578];[764-764];[1163-1183]
155	[1-486];[1095-1500]	[487-1094];[1501-1545]
156	[1-488];[740-797];[799-884];[895-974]	[489-739];[798-798];[885-894];[975-1068]
157	[1-161];[163-565];[567-701]	[162-162];[566-566];[702-1097]
158	[1-496];[692-754]	[497-691];[755-894]
159	[1-483]	[484-703]
160	[1-494]	[495-849]
161	[1-491]	[492-846]
162	[1-505];[575-759];[761-1164]	[506-574];[760-760];[1165-1176]
163	[1-699]	[700-1084]
164	[38-483];[485-556]	[1-37];[484-484];[557-1793]

165	[1-426];[1303-1444];[1717-1755];[1787-1825]	[427-1302];[1445-1716];[1756-1786];[1826-1849]
166	[2-264];[266-446];[448-519]	[1-1];[265-265];[447-447];[520-1748]
167	[1-519];[523-552]	[520-522];[553-1275]
168	[1-457];[466-571]	[458-465];[572-1023]
169	[1-54];[57-501]	[55-56];[502-1085]
170	[1-541]	[542-776]
171	[1-489]	[490-1219]
172	[1-538];[977-1468]	[539-976];[1469-1487]
173	[1-631]	[632-1915]
174	[21-776];[888-967];[969-1061];[1063-1137];[1819-1967]	[1-20];[777-887];[968-968];[1062-1062];[1138-1818];[1968-1990]
175	[1-508]	[509-1971]
176	[1-127];[129-538];[979-1470]	[128-128];[539-978];[1471-1613]
177	[1-535];[973-1173];[1177-1330];[1332-1361]	[536-972];[1174-1176];[1331-1331]
178	[1-599];[626-830];[1082-1113]	[600-625];[831-1081]
179	[1-623];[1377-1406]	[624-1376];[1407-1960]
180	[1-414];[418-464]	[415-417];[465-1443]
181	[1-522];[533-587]	[523-532];[588-605]
182	[1-78];[99-131];[136-327];[1153-1184];[1210-1274];[1284-1319];[1385-1416]	[79-98];[132-135];[328-1152];[1185-1209];[1275-1283];[1320-1384];[1417-1724]
183	[1-512];[617-805];[871-952];[1387-1422];[1621-1661]	[513-616];[806-870];[953-1386];[1423-1620];[1662-1686]
184	[1-453]	[454-463]
185	[1-773]	None
186	[1-413];[423-604];[606-739]	[414-422];[605-605];[740-753]
187	[1-117];[119-401]	[118-118];[402-754]
188	[1-511];[684-870];[872-928];[935-981]	[512-683];[871-871];[929-934];[982-998]
189	[1-605]	None
190	[2-475]	[1-1];[476-526]
191	[1-910]	None
192	[1-101];[103-668]	[102-102]
193	[1-520];[583-637]	[521-582]
194	[1-706]	None
195	[1-145];[150-451];[466-670]	[146-149];[452-465]
196	[1-509]	[510-510]
197	[1-500]	None
198	[1-503];[505-585]	[504-504];[586-667]
199	[1-498]	[499-514]
200	[1-462]	None
201	[1-551]	None
202	[1-482];[484-550]	[483-483]
203	[1-408]	None
204	[1-519];[521-649]	[520-520];[650-665]
205	[1-261];[263-415];[417-640];[642-782]	[262-262];[416-416];[641-641];[783-1008]
206	[1-455]	None

207	[1-402];[410-526]	[403-409];[527-749]
208	[1-520]	[521-594]
209	[1-197];[200-472]	[198-199];[473-2098]
210	[1-311];[314-427]	[312-313];[428-428]
211	[1-689];[735-769]	[690-734]
212	[1-517]	[518-914]
213	[2-576];[756-795];[1390-1441]	[1-1];[577-755];[796-1389];[1442-1489]
214	[1-482]	[483-776]
215	[1-498]	[499-1412]
216	[1-505];[1000-1293];[1295-1408];[1744-1773]	[506-999];[1294-1294];[1409-1743]
217	[1-102];[104-291];[293-467];[486-708];[723-831];[833-900];[910-1031];[1054-1090];[1097-1153]	[103-103];[292-292];[468-485];[709-722];[832-832];[901-909];[1032-1053];[1091-1096];[1154-1251]
218	[1-452]	[453-894]
219	[1-554];[556-598]	[555-555];[599-910]
220	[1-38];[41-95];[98-386];[388-487]	[39-40];[96-97];[387-387];[488-519]
221	[1-34];[38-220];[222-335];[337-518]	[35-37];[221-221];[336-336];[519-632]
222	[1-468]	[469-652]
223	[1-466]	[467-650]
224	[1-466]	[467-502]
225	[1-489];[653-1008]	[490-652];[1009-1739]
226	[1-657]	None
227	[1-480]	[481-888]
228	[1-501]	[502-716]
229	[1-612]	[613-654]
230	[1-477];[485-538]	[478-484];[539-635]
231	[1-476];[484-537]	[477-483];[538-634]
232	[1-367];[371-512]	[368-370];[513-583]
233	[1-305];[307-442];[460-503];[553-646]	[306-306];[443-459];[504-552];[647-753]
234	[1-260];[262-345];[347-454];[473-515];[565-658]	[261-261];[346-346];[455-472];[516-564];[659-762]
235	[1-427]	[428-537]
236	[1-465]	[466-994]
237	[1-471];[496-526];[557-587];[597-637]	[472-495];[527-556];[588-596];[638-662]
238	[1-338];[352-497]	[339-351];[498-1829]
239	[1-501]	[502-1083]
240	[1-515];[1527-1583];[1585-1687];[1692-1831]	[516-1526];[1584-1584];[1688-1691]
241	[1-515];[1526-1582];[1584-1686];[1691-1830]	[516-1525];[1583-1583];[1687-1690]

Table Vb

Seq Id No	Preferentially excluded fragments	Preferentially included fragments
1	[1-540];[556-615];[2061-2096];[2098-2201]	[541-555];[616-2060];[2097-2097]
2	[1-511];[533-619];[621-690];[730-1132]	[512-532];[620-620];[691-729];[1133-1631]
3	[2-539];[1178-1245]	[1-1];[540-1177]
4	[1-250];[297-383];[386-514];[1025-1064]	[251-296];[384-385];[515-1024];[1065-1623]
5	[27-116];[118-391]	[1-26];[117-117];[392-1454]
6	[1-93];[96-168];[170-262];[264-461]	[94-95];[169-169];[263-263];[462-1639]
7	[1-95];[97-451]	[96-96];[452-1768]
8	[1-502];[1314-1491]	[503-1313];[1492-1510]
9	[1-864]	[865-882]
10	[1-428]	[429-1849]
11	[1-454];[482-514]	[455-481];[515-565]
12	[1-375];[379-511];[533-690];[730-783];[814-1164]	[376-378];[512-532];[691-729];[784-813];[1165-1663]
13	[2-337];[339-556]	[1-1];[338-338];[557-744]
14	[29-366];[368-507]	[1-28];[367-367];[508-1759]
15	[29-366];[368-524]	[1-28];[367-367];[525-1755]
16	[1-641]	[642-936]
17	[1-708];[711-747]	[709-710]
18	[1-639]	[640-1884]
19	[1-631]	[632-691]
20	[3-416];[418-490]	[1-2];[417-417];[491-1138]
21	[1-468]	None
22	[1-720]	None
23	[1-711]	[712-727]
24	[1-469]	[470-470]
25	[1-231];[234-488]	[232-233];[489-987]
26	[1-296];[300-642];[644-737]	[297-299];[643-643];[738-908]
27	[1-306];[308-762]	[307-307]
28	[1-446];[448-1102]	[447-447]
29	[1-436]	None
30	[7-334];[1420-1468];[1474-1614];[1616-1804];[1845-1919]	[1-6];[335-1419];[1469-1473];[1615-1615];[1805-1844];[1920-1938]
31	[1-342];[345-519];[823-893];[977-1016]	[343-344];[520-822];[894-976];[1017-1116]
32	[1-517];[821-891];[975-1014]	[518-820];[892-974];[1015-1114]
33	[36-352];[354-457];[728-832];[834-1096];[1253-1289];[1291-1350];[1352-1412];[1726-1873]	[1-35];[353-353];[458-727];[833-833];[1097-1252];[1290-1290];[1351-1351];[1413-1725];[1874-2072]
34	[1-409]	None
35	[14-105]	[1-13];[106-836]
36	[1-572];[1120-1271]	[573-1119];[1272-1323]
37	[20-98];[100-510];[1591-1681];[1683-1870]	[1-19];[99-99];[511-1590];[1682-1682];[1871-1945]
38	[1-547]	[548-1330]

39	[1-445]	[446-2124]
40	[1-473];[475-528]	[474-474];[529-1159]
41	[16-506];[1587-1866]	[1-15];[507-1586];[1867-1953]
42	[2-234];[244-451];[974-1226]	[1-1];[235-243];[452-973];[1227-1688]
43	[1-455];[1670-1925]	[456-1669];[1926-1942]
44	[1-579];[815-1031]	[580-814];[1032-1657]
45	[1-489];[1012-1264]	[490-1011];[1265-1733]
46	[1-400];[1184-1223];[1225-1705];[1740-1818]	[401-1183];[1224-1224];[1706-1739];[1819-1871]
47	[1-529];[1326-1505]	[530-1325];[1506-1523]
48	[1-131];[133-510];[560-589]	[132-132];[511-559];[590-832]
49	[1-130];[132-509];[559-588]	[131-131];[510-558];[589-831]
50	[1-650];[652-868];[873-913]	[651-651];[869-872];[914-917]
51	[1-504];[515-605]	[505-514];[606-621]
52	[1-535]	[536-673]
53	[2-563]	[1-1];[564-897]
54	[1-527];[802-870];[882-934];[966-1018];[1037-1080]	[528-801];[871-881];[935-965];[1019-1036];[1081-1101]
55	[1-326];[328-505]	[327-327];[506-1047]
56	[1-340]	[341-925]
57	[1-528]	[529-1240]
58	[1-108];[115-151];[154-340];[342-529]	[109-114];[152-153];[341-341];[530-902]
59	[4-485];[1566-1656];[1658-1845]	[1-3];[486-1565];[1657-1657];[1846-1969]
60	[1-283]	[284-1132]
61	[9-468]	[1-8];[469-631]
62	[1-525];[689-722]	[526-688]
63	[1-88];[90-192];[194-265];[296-409]	[89-89];[193-193];[266-295];[410-1442]
64	[1-517]	[518-795]
65	[1-406];[408-739]	[407-407];[740-1236]
66	[1-489];[849-881]	[490-848]
67	[1-505]	[506-524]
68	[1-325];[328-441];[444-504]	[326-327];[442-443];[505-1472]
69	[1-524];[636-715];[717-809];[811-885];[1567-1715]	[525-635];[716-716];[810-810];[886-1566];[1716-1737]
70	[12-487]	[1-11];[488-1637]
71	[12-487]	[1-11];[488-1636]
72	[1-451]	[452-1758]
73	[1-167];[242-464]	[168-241];[465-1647]
74	[1-167];[242-464]	[168-241];[465-1646]
75	[1-471]	[472-1963]
76	[1-358];[360-543];[655-734];[736-828];[830-904];[1586-1734]	[359-359];[544-654];[735-735];[829-829];[905-1585];[1735-1757]
77	[3-34];[36-474];[582-770];[1709-1746];[1748-1785];[1825-1899]	[1-2];[35-35];[475-581];[771-1708];[1747-1747];[1786-1824];[1900-2027]
78	[1-75];[77-319];[914-1052];[1063-1126];[1168-1203]	[76-76];[320-913];[1053-1062];[1127-1167];[1204-1880]
79	[1-425]	[426-584]

80	[1-752];[947-1017];[1084-1170]	[753-946];[1018-1083];[1171-1351]
81	[1-496];[498-720]	[497-497]
82	[1-324]	[325-1029]
83	[1-477];[1474-1529];[1537-1566];[1577-1616];[1622-1662];[1717-1753]	[478-1473];[1530-1536];[1567-1576];[1617-1621];[1663-1716];[1754-1788]
84	[1-496];[499-568];[752-805]	[497-498];[569-751]
85	[1-527]	[528-814]
86	[1-360]	[361-598]
87	[1-78];[80-583];[625-699]	[79-79];[584-624]
88	[1-889]	[890-905]
89	[1-513]	[514-514]
90	[1-122];[124-155];[157-435];[437-517]	[123-123];[156-156];[436-436];[518-518]
91	[1-133];[165-808]	[134-164]
92	[1-725]	[726-737]
93	[1-409]	[410-728]
94	[1-331]	[332-582]
95	[1-410]	[411-1913]
96	[1-501]	[502-670]
97	[1-141];[143-431]	[142-142];[432-939]
98	[1-193]	[194-661]
99	[1-629]	[630-647]
100	[1-520];[862-954];[976-1005]	[521-861];[955-975];[1006-1006]
101	[1-489];[581-961];[1010-1059]	[490-580];[962-1009]
102	[1-485]	[486-514]
103	[1-540]	[541-1158]
104	[1-556]	[557-1563]
105	[1-868];[870-1006]	[869-869];[1007-1621]
106	[1-491]	[492-557]
107	[1-573]	[574-600]
108	[1-457];[586-1110]	[458-585];[1111-1129]
109	[1-521];[655-778]	[522-654]
110	[1-416];[478-614];[616-990];[992-1065];[1068-1283]	[417-477];[615-615];[991-991];[1066-1067];[1284-1301]
111	[1-416];[478-614];[628-989];[991-1064];[1067-1282]	[417-477];[615-627];[990-990];[1065-1066];[1283-1300]
112	[2-429];[1161-1202];[1212-1388];[1392-1589]	[1-1];[430-1160];[1203-1211];[1389-1391];[1590-1617]
113	[1-487]	[488-1634]
114	[1-70];[86-496]	[71-85];[497-693]
115	[1-358];[360-558]	[359-359];[559-784]
116	[1-215];[218-495];[527-607]	[216-217];[496-526];[608-804]
117	[1-466]	[467-484]
118	[1-515];[906-963]	[516-905];[964-985]
119	[1-744];[746-816]	[745-745];[817-839]
120	[1-85];[87-521]	[86-86];[522-583]
121	[1-532]	[533-1024]

122	[1-318];[325-517];[567-660]	[319-324];[518-566];[661-760]
123	[1-498]	[499-594]
124	[1-427]	[428-559]
125	[1-642]	[643-744]
126	[1-341];[350-696]	[342-349];[697-824]
127	[1-482]	[483-526]
128	[1-338]	[339-618]
129	[1-191];[193-429];[450-678]	[192-192];[430-449];[679-776]
130	[19-463];[465-544]	[1-18];[464-464];[545-998]
131	[1-470]	[471-779]
132	[1-533]	[534-1025]
133	[1-498]	[499-607]
134	[1-168];[170-326];[328-471];[552-738]	[169-169];[327-327];[472-551];[739-774]
135	[1-346];[348-395];[440-473]	[347-347];[396-439];[474-611]
136	[1-324];[343-436]	[325-342];[437-925]
137	[1-186];[188-251];[255-517]	[187-187];[252-254];[518-674]
138	[1-488]	[489-1725]
139	[1-101];[103-190];[292-327];[1091-1161];[1228-1314]	[102-102];[191-291];[328-1090];[1162-1227];[1315-1474]
140	[1-465];[516-653]	[466-515]
141	[1-761];[763-857];[912-1326]	[762-762];[858-911];[1327-1490]
142	[1-476]	[477-661]
143	[1-531];[1471-1508];[1510-1547];[1587-1661]	[532-1470];[1509-1509];[1548-1586];[1662-1789]
144	[1-492];[503-536]	[493-502];[537-2006]
145	[1-570]	[571-1096]
146	[1-536];[621-703];[729-1075];[1198-1445]	[537-620];[704-728];[1076-1197];[1446-1666]
147	[1-555];[578-628]	[556-577];[629-1687]
148	[1-444];[1201-1474];[1480-1516]	[445-1200];[1475-1479];[1517-1747]
149	[1-613];[626-658]	[614-625]
150	[4-199];[201-419];[421-492]	[1-3];[200-200];[420-420];[493-2045]
151	[1-509]	[510-788]
152	[1-483];[485-578]	[484-484];[579-1931]
153	[1-497]	[498-514]
154	[5-509];[579-763];[765-1162]	[1-4];[510-578];[764-764];[1163-1183]
155	[1-486];[1095-1500]	[487-1094];[1501-1545]
156	[1-488];[740-797];[799-884];[895-974]	[489-739];[798-798];[885-894];[975-1068]
157	[1-161];[163-565];[567-701]	[162-162];[566-566];[702-1097]
158	[1-496];[692-754]	[497-691];[755-894]
159	[1-483]	[484-703]
160	[1-494]	[495-849]
161	[1-491]	[492-846]
162	[1-505];[575-759];[761-1164]	[506-574];[760-760];[1165-1176]
163	[1-699]	[700-1084]
164	[38-483];[485-556]	[1-37];[484-484];[557-1793]
165	[1-426];[1303-1444];[1717-1755];[1787-1825]	[427-1302];[1445-1716];[1756-1786];[1826-1849]

166	[2-264];[266-446];[448-519]	[1-1];[265-265];[447-447];[520-1748]
167	[1-519];[523-552]	[520-522];[553-1275]
168	[1-457];[466-571]	[458-465];[572-1023]
169	[1-54];[57-501]	[55-56];[502-1085]
170	[1-541]	[542-776]
171	[1-489]	[490-1219]
172	[1-538];[977-1468]	[539-976];[1469-1487]
173	[1-631]	[632-1915]
174	[21-776];[888-967];[969-1061];[1063-1137];[1819-1967]	[1-20];[777-887];[968-968];[1062-1062];[1138-1818];[1968-1990]
175	[1-508]	[509-1971]
176	[1-127];[129-538];[979-1470]	[128-128];[539-978];[1471-1613]
177	[1-535];[973-1173];[1177-1330];[1332-1361]	[536-972];[1174-1176];[1331-1331]
178	[1-599];[626-830];[1082-1113]	[600-625];[831-1081]
179	[1-623];[1377-1406]	[624-1376];[1407-1960]
180	[1-414];[418-464]	[415-417];[465-1443]
181	[1-522];[533-587]	[523-532];[588-605]
182	[1-78];[99-131];[136-327];[1153-1184];[1210-1274];[1284-1319];[1385-1416]	[79-98];[132-135];[328-1152];[1185-1209];[1275-1283];[1320-1384];[1417-1724]
183	[1-512];[617-805];[871-952];[1387-1422];[1621-1661]	[513-616];[806-870];[953-1386];[1423-1620];[1662-1686]
184	[1-453]	[454-463]
185	[1-773]	None
186	[1-413];[423-604];[606-739]	[414-422];[605-605];[740-753]
187	[1-117];[119-401]	[118-118];[402-754]
188	[1-511];[684-870];[872-928];[935-981]	[512-683];[871-871];[929-934];[982-998]
189	[1-605]	None
190	[2-475]	[1-1];[476-526]
191	[1-910]	None
192	[1-101];[103-668]	[102-102]
193	[1-520];[583-637]	[521-582]
194	[1-706]	None
195	[1-145];[150-451];[466-670]	[146-149];[452-465]
196	[1-509]	[510-510]
197	[1-500]	None
198	[1-503];[505-585]	[504-504];[586-667]
199	[1-498]	[499-514]
200	[1-462]	None
201	[1-551]	None
202	[1-482];[484-550]	[483-483]
203	[1-408]	None
204	[1-519];[521-649]	[520-520];[650-665]
205	[1-261];[263-415];[417-640];[642-782]	[262-262];[416-416];[641-641];[783-1008]
206	[1-455]	None
207	[1-402];[410-526]	[403-409];[527-749]
208	[1-520]	[521-594]

209	[1-197];[200-472]	[198-199];[473-2098]
210	[1-311];[314-427]	[312-313];[428-428]
211	[1-689];[735-769]	[690-734]
212	[1-517]	[518-914]
213	[2-576];[756-795];[1390-1441]	[1-1];[577-755];[796-1389];[1442-1489]
214	[1-482]	[483-776]
215	[1-498]	[499-1412]
216	[1-505];[1000-1293];[1295-1408];[1744-1773]	[506-999];[1294-1294];[1409-1743]
217	[1-102];[104-291];[293-467];[486-708];[723-831];[833-900];[910-1031];[1054-1090];[1097-1153]	[103-103];[292-292];[468-485];[709-722];[832-832];[901-909];[1032-1053];[1091-1096];[1154-1251]
218	[1-452]	[453-894]
219	[1-554];[556-598]	[555-555];[599-910]
220	[1-38];[41-95];[98-386];[388-487]	[39-40];[96-97];[387-387];[488-519]
221	[1-34];[38-220];[222-335];[337-518]	[35-37];[221-221];[336-336];[519-632]
222	[1-468]	[469-652]
223	[1-466]	[467-650]
224	[1-466]	[467-502]
225	[1-489];[653-1008]	[490-652];[1009-1739]
226	[1-657]	None
227	[1-480]	[481-888]
228	[1-501]	[502-716]
229	[1-612]	[613-654]
230	[1-477];[485-538]	[478-484];[539-635]
231	[1-476];[484-537]	[477-483];[538-634]
232	[1-367];[371-512]	[368-370];[513-583]
233	[1-305];[307-442];[460-503];[553-646]	[306-306];[443-459];[504-552];[647-753]
234	[1-260];[262-345];[347-454];[473-515];[565-658]	[261-261];[346-346];[455-472];[516-564];[659-762]
235	[1-427]	[428-537]
236	[1-465]	[466-994]
237	[1-471];[496-526];[557-587];[597-637]	[472-495];[527-556];[588-596];[638-662]
238	[1-338];[352-497]	[339-351];[498-1829]
239	[1-501]	[502-1083]
240	[1-515];[1527-1583];[1585-1687];[1692-1831]	[516-1526];[1584-1584];[1688-1691]
241	[1-515];[1526-1582];[1584-1686];[1691-1830]	[516-1525];[1583-1583];[1687-1690]

Table VI

Seq Id No	Designation of domain	Database	Positions f domains
242	Cell attachment sequence	PROSITE	141-143
242	Peptidase family M20/M25/M40	PFAM	107-451
244	Mitochondrial energy transfer proteins signature	PROSITE	26-35
244	Mitochondrial energy transfer proteins signature	PROSITE	199-208
244	Mitochondrial carrier proteins	PFAM	5-84;87-175;178-272
244	Mitochondrial energy transfer proteins.	BLOCKSPLUS	12-36
244	Mitochondrial energy transfer proteins.	BLOCKSPLUS	13-36
244	Mitochondrial energy transfer proteins.	BLOCKSPLUS	131-144
245	Leucine zipper pattern	PROSITE	371-392
249	Leucine zipper pattern	PROSITE	20-41
251	Mitochondrial energy transfer proteins signature	PROSITE	26-35
251	Mitochondrial carrier proteins	PFAM	5-72
251	Mitochondrial energy transfer proteins.	BLOCKSPLUS	12-36
251	Mitochondrial energy transfer proteins.	BLOCKSPLUS	13-36
254	Pancreatic ribonuclease family signature	PROSITE	63-69
254	Pancreatic ribonucleases	PFAM	26-143
254	PANCREATIC RIBONUCLEASE FAMILY SIGNATURE	BLOCKSPLUS	49-69
254	Pancreatic ribonuclease family proteins.	BLOCKSPLUS	115-140
254	PANCREATIC RIBONUCLEASE FAMILY SIGNATURE	BLOCKSPLUS	92-110
254	PANCREATIC RIBONUCLEASE FAMILY SIGNATURE	BLOCKSPLUS	114-133
254	Pancreatic ribonuclease family proteins.	BLOCKSPLUS	30-40
254	PANCREATIC RIBONUCLEASE FAMILY SIGNATURE	BLOCKSPLUS	114-137
254	PANCREATIC RIBONUCLEASE FAMILY SIGNATURE	BLOCKSPLUS	69-86
255	L-lactate dehydrogenase active site	PROSITE	239-245
255	lactate/malate dehydrogenase	PFAM	71-380
255	L-lactate dehydrogenase proteins.	BLOCKSPLUS	186-224
255	L-LACTATE DEHYDROGENASE SIGNATURE	BLOCKSPLUS	96-121
255	L-lactate dehydrogenase proteins.	BLOCKSPLUS	71-102
255	L-lactate dehydrogenase proteins.	BLOCKSPLUS	238-256
255	L-LACTATE DEHYDROGENASE SIGNATURE	BLOCKSPLUS	183-203
255	L-lactate dehydrogenase proteins.	BLOCKSPLUS	288-323

255	L-LACTATE DEHYDROGENASE SIGNATURE	BLOCKSPLUS	207-224
255	L-LACTATE DEHYDROGENASE SIGNATURE	BLOCKSPLUS	71-92
255	L-lactate dehydrogenase proteins.	BLOCKSPLUS	138-167
256	lactate/malate dehydrogenase	PFAM	71-124
256	L-LACTATE DEHYDROGENASE SIGNATURE	BLOCKSPLUS	96-121
256	L-lactate dehydrogenase proteins.	BLOCKSPLUS	71-102
256	L-LACTATE DEHYDROGENASE SIGNATURE	BLOCKSPLUS	71-92
256	L-lactate dehydrogenase proteins.	BLOCKSPLUS	71-100
256	L-LACTATE DEHYDROGENASE SIGNATURE	BLOCKSPLUS	71-84
257	Leucine zipper pattern	PROSITE	155-176
259	HORMA domain	PFAM	22-230
261	Leucine zipper pattern	PROSITE	142-163
261	Leucine zipper pattern	PROSITE	170-191
263	Leucine zipper pattern	PROSITE	15-36
264	Ubiquitin family	PFAM	1-82
264	Ubiquitin domain proteins.	BLOCKSPLUS	17-62
264	Ubiquitin domain proteins.	BLOCKSPLUS	21-68
264	Ubiquitin domain proteins.	BLOCKSPLUS	26-68
264	Ubiquitin domain proteins.	BLOCKSPLUS	17-68
266	u-PAR/Ly-6 domain	PFAM	60-119
266	Squash family of serine protease inhibit	PFAM	32-47
267	Zinc finger, C2H2 type, domain proteins.	BLOCKSPLUS	185-202
271	LBP / BPI / CETP family signature	PROSITE	28-60
271	Pyrokinins signature	PROSITE	324-328
271	LBP / BPI / CETP family	PFAM	10-479
271	LBP / BPI / CETP family proteins.	BLOCKSPLUS	72-118
271	LBP / BPI / CETP family proteins.	BLOCKSPLUS	209-253
271	LBP / BPI / CETP family proteins.	BLOCKSPLUS	28-58
271	LBP / BPI / CETP family proteins.	BLOCKSPLUS	275-309
271	LBP / BPI / CETP family proteins.	BLOCKSPLUS	76-113
272	Zinc finger, C3HC4 type (RING finger), signature	PROSITE	102-111
272	Zinc finger, C3HC4 type (RING finger)	PFAM	87-129
272	Zinc finger, C3HC4 type (RING finger), proteins.	BLOCKSPLUS	102-111
273	Zinc finger, C3HC4 type (RING finger), signature	PROSITE	30-39
273	Zinc finger, C3HC4 type (RING finger)	PFAM	15-57
273	Zinc finger, C3HC4 type (RING finger), proteins.	BLOCKSPLUS	30-39
274	RNA 3'-terminal phosphate cyclase signature	PROSITE	157-167
274	RNA 3'-terminal phosphate cyclase	PFAM	1-368

274	RNA 3'-terminal phosphate cyclase proteins.	BLOCKSPLUS	12-44
274	RNA 3'-terminal phosphate cyclase proteins.	BLOCKSPLUS	157-168
275	Ribosomal L27 protein	PFAM	31-86
277	Cell attachment sequence	PROSITE	292-294
277	DHHC zinc finger domain	PFAM	140-204
279	Endogenous opioids neuropeptides precursors signature	PROSITE	26-65
279	Vertebrate endogenous opioids neurope	PFAM	3-257
279	Endogenous opioids neuropeptides precursors proteins.	BLOCKSPLUS	100-126
279	Endogenous opioids neuropeptides precursors proteins.	BLOCKSPLUS	209-237
279	Endogenous opioids neuropeptides precursors proteins.	BLOCKSPLUS	43-66
279	Endogenous opioids neuropeptides precursors proteins.	BLOCKSPLUS	18-38
279	Endogenous opioids neuropeptides precursors proteins.	BLOCKSPLUS	24-36
279	Endogenous opioids neuropeptides precursors proteins.	BLOCKSPLUS	105-125
280	Leucine zipper pattern	PROSITE	136-157
280	Leucine zipper pattern	PROSITE	272-293
283	Immunoglobulins and major histocompatibility complex proteins signature	PROSITE	380-386
283	Immunoglobulin domain	PFAM	205-285;318-384
283	Immunoglobulins and major histocompatibility complex proteins.	BLOCKSPLUS	319-336
284	Fucosyl transferase	PFAM	70-406
285	FAD/NAD-binding Cytochrome reductase	PFAM	27-149
285	Oxidoreductase FAD/NAD-binding domain	PFAM	176-290
285	Eukaryotic molybdopterin oxidoreductases proteins.	BLOCKSPLUS	58-86
285	CYTOCHROME B5 REDUCTASE SIGNATURE	BLOCKSPLUS	75-86
285	CYTOCHROME B5 REDUCTASE SIGNATURE	BLOCKSPLUS	274-283
285	CYTOCHROME B5 REDUCTASE SIGNATURE	BLOCKSPLUS	141-156
285	Eukaryotic molybdopterin oxidoreductases proteins.	BLOCKSPLUS	274-286
285	Eukaryotic molybdopterin oxidoreductases proteins.	BLOCKSPLUS	60-85
285	CYTOCHROME B5 REDUCTASE SIGNATURE	BLOCKSPLUS	181-198
285	FLAVOPROTEIN PYRIDINE NUCLEOTIDE CYTOCHROME REDUCTASE SIGNATURE	BLOCKSPLUS	181-197

286	Immunoglobulins and major histocompatibility complex proteins signature	PROSITE	380-386
286	Immunoglobulin domain	PFAM	205-285;318-384
286	Immunoglobulins and major histocompatibility complex proteins.	BLOCKSPLUS	319-336
287	Leucine zipper pattern	PROSITE	126-147
288	Leucine zipper pattern	PROSITE	20-41
291	Tissue inhibitors of metalloproteinases signature	PROSITE	24-36
291	Tissue inhibitor of metalloproteinases	PFAM	22-199
291	Tissue inhibitors of metalloproteinases proteins.	BLOCKSPLUS	21-46
291	Tissue inhibitors of metalloproteinases proteins.	BLOCKSPLUS	106-148
291	Tissue inhibitors of metalloproteinases proteins.	BLOCKSPLUS	81-95
291	Tissue inhibitors of metalloproteinases proteins.	BLOCKSPLUS	61-72
294	Domain of unknown function DUF59	PFAM	31-135
296	Immunoglobulin domain	PFAM	141-197
297	TonB-dependent receptor proteins signature 1	PROSITE	1-42
298	Fibroblast growth factor	PFAM	48-129
299	Bola-like protein	PFAM	39-114
299	PROTEIN BOLA TRANSCRIPTION REGULATION AC.	BLOCKSPLUS	68-98
301	Cell attachment sequence	PROSITE	172-174
303	Ribosomal L27 protein	PFAM	31-115
304	Leucine rich repeat C-terminal domain	PFAM	173-222
304	Leucine Rich Repeat	PFAM	92-115;116-139;140-163;164-185
309	Leucine rich repeat C-terminal domain	PFAM	173-222
309	Leucine Rich Repeat	PFAM	92-115;116-139;140-163;164-185
311	NOL1/NOP2/sun family	PFAM	201-276;353-378
311	NOL1/NOP2/sun family proteins.	BLOCKSPLUS	230-245
311	NOL1/NOP2/sun family proteins.	BLOCKSPLUS	231-245
312	NOL1/NOP2/sun family	PFAM	201-276
312	NOL1/NOP2/sun family proteins.	BLOCKSPLUS	230-245
312	NOL1/NOP2/sun family proteins.	BLOCKSPLUS	231-245
314	Leucine zipper pattern	PROSITE	8-29
315	Leucine zipper pattern	PROSITE	8-29
341	Immunoglobulin domain	PFAM	45-112
349	CDP-alcohol phosphatidyltransferases signature	PROSITE	54-76
349	Cytochrome b/b6 Qo site signature	PROSITE	97-102
354	SAM domain (Sterile alpha motif)	PFAM	82-147
361	Ribosomal Proteins L2	PFAM	96-124
368	DAD family	PFAM	1-78
370	Ribosomal protein L34	PFAM	51-92

385	Kelch motif	PFAM	20-66;68-114;116-162;164-209;211-265;270-316
386	SPRY domain	PFAM	85-205
388	PHD-finger.	BLOCKSPLUS	329-339
389	Eukaryotic thiol (cysteine) proteases histidine active site	PROSITE	268-278
389	Heat shock hsp70 proteins family signature 3	PROSITE	332-346
389	Hsp70 protein	PFAM	3-509
390	Eukaryotic-type carbonic anhydrase	PFAM	20-59
391	PMP-22/EMP/MP20/Claudin family	PFAM	4-162
392	Sec1 family.	BLOCKSPLUS	89-107
393	PMP-22/EMP/MP20/Claudin family	PFAM	4-182
394	Myc-type, 'helix-loop-helix' dimerization domain signature	PROSITE	13-28
395	Glutathione S-transferases.	PFAM	47-122;260-309
396	Transmembrane 4 family signature	PROSITE	112-134
396	Transmembrane 4 family	PFAM	66-273
396	Transmembrane 4 family proteins.	BLOCKSPLUS	108-146
396	TRANSMEMBRANE FOUR FAMILY SIGNATURE	BLOCKSPLUS	129-151
396	TRANSMEMBRANE FOUR FAMILY SIGNATURE	BLOCKSPLUS	108-127
396	TRANSMEMBRANE FOUR FAMILY SIGNATURE	BLOCKSPLUS	247-274
396	TRANSMEMBRANE FOUR FAMILY SIGNATURE	BLOCKSPLUS	129-150
396	TRANSMEMBRANE FOUR FAMILY SIGNATURE	BLOCKSPLUS	128-154
397	ATP/GTP-binding site motif A (P-loop)	PROSITE	6-13
397	ADP-ribosylation factor family	PFAM	2-172
398	Isochorismatase family	PFAM	17-147
399	PAP2 superfamily	PFAM	19-175
400	Zinc carboxypeptidases, zinc-binding region 2 signature	PROSITE	117-127
401	Zinc finger, C2H2 type, domain	PROSITE	36-57
401	Zinc finger, C2H2 type, domain	PROSITE	73-93
401	Zinc finger, C2H2 type, domain	PROSITE	114-134
401	Zinc finger, C2H2 type, domain	PROSITE	145-165
401	Zinc finger, C2H2 type	PFAM	34-57;71-93;112-134;143-165
401	Zinc finger, C2H2 type, domain proteins.	BLOCKSPLUS	145-162
401	Zinc finger, C2H2 type, domain proteins.	BLOCKSPLUS	114-131
401	Zinc finger, C2H2 type, domain proteins.	BLOCKSPLUS	73-90
402	Zinc finger, C2H2 type, domain	PROSITE	113-133

402	Zinc finger, C2H2 type, domain	PROSITE	144-164
402	Regulator of chromosome condensation (RCC1) signature 2	PROSITE	65-75
402	Zinc finger, C2H2 type	PFAM	111-133;142-164
402	Zinc finger, C2H2 type, domain proteins.	BLOCKSPLUS	144-161
402	Zinc finger, C2H2 type, domain proteins.	BLOCKSPLUS	113-130
403	Glutathione S-transferases.	PFAM	47-122;260-309
405	PMP-22/EMP/MP20/Claudin family	PFAM	4-182
406	WD domain, G-beta repeat	PFAM	267-304;333-370
408	Rhomboid family	PFAM	186-323
410	Ank repeat	PFAM	47-79
410	REPEAT PROTEIN ANK NUCLEAR ANKYR.	BLOCKSPLUS	78-89
410	Ank repeat proteins.	BLOCKSPLUS	48-56
412	Serine proteases, subtilase family, aspartic acid proteins.	BLOCKSPLUS	165-178
414	Sir2 family	PFAM	84-268
416	Kelch motif	PFAM	20-66;68-114;116-162;164-209;211-265;270-316
418	Zinc-binding dehydrogenases	PFAM	16-313
426	Leucine zipper pattern	PROSITE	144-165
447	Cytochrome c family heme-binding site signature	PROSITE	19-24
447	Immunoglobulins and major histocompatibility complex proteins signature	PROSITE	17-23
453	eIF-6 family	PFAM	3-103
454	Cell attachment sequence	PROSITE	226-228
456	Leucine zipper pattern	PROSITE	211-232
457	Leucine zipper pattern	PROSITE	236-257
466	Zinc finger, C3HC4 type (RING finger), signature	PROSITE	56-65
466	SPRY domain	PFAM	375-500
466	Zinc finger, C3HC4 type (RING finger)	PFAM	41-81
466	B-box zinc finger.	PFAM	110-153
466	Domain in SPla and the RYanodine Receptor.	BLOCKSPLUS	359-381
466	Domain in SPla and the RYanodine Receptor.	BLOCKSPLUS	443-457
466	Domain in SPla and the RYanodine Receptor.	BLOCKSPLUS	359-380
466	Zinc finger, C3HC4 type (RING finger), proteins.	BLOCKSPLUS	56-65
479	UBX domain	PFAM	329-408
481	TBC domain	PFAM	65-171
481	Probable rabGAP domain proteins.	BLOCKSPLUS	153-159
482	TBC domain	PFAM	65-177

482	Probable rabGAP domain proteins.	BLOCKSPLUS	153-159
-----	----------------------------------	------------	---------

Table VII

Seq Id No	Epitopes
242	98..109;119..127;136..147;156..170;242..248;255..265;318..328;356..363;399..407;443..450;475..490
243	3..9;59..65;69..79;113..126;142..155;193..198;212..220;231..245;302..315
244	29..36;33..42;79..87;139..147;269..274
245	101..107;141..151;156..165;196..207;225..233;242..251;253..260;284..298;323..330;339..347;395..406
247	41..51;108..120;121..131;190..200;255..261;302..307
248	5..11;38..46;52..60;75..83;92..99;133..150;167..183;187..200;210..219;244..252;270..286;335..345;354..371;390..397
249	68..80;91..99;132..138;185..193;265..273;276..293;295..306;305..329;327..341;347..358;394..403
250	28..37;60..67;73..81
251	33..45;64..71
252	20..30;35..45;49..59;74..83
253	3..9;59..65
254	22..33;35..52;53..67;70..77;80..100;106..117;142..147
255	116..123;147..156;201..208;262..278
256	10..15;116..121
257	41..51;52..66;72..80;94..101;120..127;134..147;180..193;204..210;227..240
258	147..157;189..199
259	52..59;66..76;103..113;115..127;131..140;143..148;181..199;242..250;253..262;262..273;279..289;330..341;342..366;373..394
260	94..107;112..119;125..134
261	121..126;144..152;213..224
263	44..50
264	51..58;82..90;153..164
266	15..20;38..49;76..81;95..105
267	74..91;94..99;117..130;140..154;153..161;175..184;201..210;228..240;250..255
268	36..42;43..54
269	41..46;64..73;80..100;106..122;160..172
270	38..48;82..88
271	34..40;72..79;111..123;146..153;251..259;307..314;316..322;372..377;436..444
272	12..17;51..58;75..85;128..136
273	4..13;56..64
274	34..46;120..127;157..163;182..191;231..240;259..267;273..279;291..299;344..355
275	30..55;72..78
276	27..35;37..45;49..61;61..77;102..109;144..152;170..180;179..188
277	61..67;147..152;154..166;284..299;308..313
278	72..82;451..461;532..541
279	24..31;72..84;83..92;97..111;144..149;161..182;181..189;192..198;204..214;216..233;241..254;256..263

280	15..20;50..56;61..66;204..212;251..260;354..362
281	24..38;44..52
282	72..82;451..461;532..541
283	1..6;21..31;77..83;115..120;228..237;276..281;335..343;401..407;440..456;456..468
284	4..9;39..47;50..66;78..94;111..122;132..141;169..174;190..202;213..220;243..252;261..274;282..300;369..376;379..389;395..403
285	29..38;42..47;58..65;100..110;121..134;156..161;161..173;201..207;230..239;243..254;290..302
286	1..6;21..28;77..83;115..120;228..237;276..281;335..343;401..407
287	2..10;94..104;248..258;268..286
288	68..80;91..99;132..138;185..194;262..270;273..288;291..301;300..324;322..336;342..353;389..398
289	23..38
291	28..35;96..104;134..144;159..167;177..187;191..198
292	1..7;56..64;66..73;77..92
293	40..45;99..109
294	47..57;120..126
295	31..61;76..82;143..149;156..169
296	133..143;151..156;161..167;169..181;185..194
297	50..58;59..69;113..123;120..137
298	45..55;52..63;106..117;118..128;126..131;148..155;157..164;172..190;212..221;232..247
299	51..59;82..87;113..125;124..135
300	72..82;451..461;532..541
301	43..52;88..105;192..211;255..271
302	3..18;37..44;57..65;70..76;98..113;121..134
303	30..55;72..77;82..88;106..113
304	2..11;33..42;48..54;55..63;122..131;147..154;168..180;200..209;211..220;226..233;268..278;286..291
305	22..31
306	5..11;25..35;72..81;124..134;147..157;163..178;177..186;185..195;207..217
307	23..38
308	66..72;84..100
309	2..11;33..42;48..54;55..63;122..131;147..154;168..180;200..209;211..220;226..233;268..278;286..291
310	45..52;60..68;88..94;99..109;113..120;121..134;162..171;169..184;194..202;209..215;223..235;239..248;273..281;292..301;319..329;336..341;389..394;398..405;421..426
311	15..21;28..35;82..91;113..120;125..133;153..167;236..243;291..298;307..312;316..327;390..396;406..413;436..457
312	15..21;28..35;82..91;113..120;125..133;153..167;236..243;291..298;307..312;316..327;352..370;370..382
313	38..46;52..60;75..83;92..99;133..150;167..183;187..200;210..219;239..256
314	36..42;52..58;65..70;80..87;143..155;161..168;176..185;203..208;263..272
315	36..42;52..58;65..70;80..87;143..155;161..168
316	33..47;49..58;106..117;125..132

317	45..53;54..68;88..94;99..109;113..120;121..134;162..171;169..184;194..202;209..215;223..235;239..248;273..281;292..301;319..329;336..341;389..394;398..405;421..426
318	42..54;69..77;124..130;148..153;156..165;186..210;219..227;271..286;293..300
319	15..21;25..37;36..59;58..64;80..89;86..102;105..119
320	1..6;81..89
321	111..116
322	1..21;50..68;76..85
323	11..16;49..68
325	14..20;40..55;69..76;122..131;128..138;158..166
326	18..45;61..68;81..89;110..141;142..149
327	43..48;53..62;85..95
328	4..9;34..46
329	1..7;58..69
330	34..42;46..52;56..66
331	59..69;72..80;80..89;90..98;110..115
332	14..24;50..56
333	1..6;41..47;86..96;120..134
334	17..23;41..47;50..57;85..90;96..105;154..159;181..192;192..198
335	4..9;21..28
336	49..65;70..80
337	20..30;36..42;53..61;74..83;110..119;125..138;137..142
338	21..55;55..65;62..81;88..100;99..107
339	33..47;52..60;73..82
340	5..12;13..21;32..51;65..72
341	19..30;44..52;51..61;68..82;96..108
342	18..26;44..54;71..77
343	21..33;57..66;68..77;80..89;91..98
344	36..44;89..95;110..116;181..193
345	19..28;34..41;43..55;80..86;119..124;159..165;167..176
346	26..35;39..49;88..95;136..145;207..217
347	2..10;26..32;52..68;75..86
348	71..77;82..89;114..125
349	54..66;70..76;294..302
350	16..25;31..37;69..80
351	68..73;118..129;135..148
352	68..73;118..129;135..148
353	10..18;61..69;68..77
354	13..20;44..51;57..85;94..102;143..151
355	18..30;38..43;43..54
356	1..6;40..45;120..126;129..154;153..162
357	49..59;133..141;152..167
358	9..26;80..85;96..102
359	98..104;181..187
360	1..6;44..54;113..123;147..161
361	64..74;77..90;112..141
362	20..30;37..42;53..61;74..83;110..119;125..134;138..147
363	1..11;26..31;53..60;97..105;110..117;141..146
364	16..24
365	7..15

367	8..14
368	16..23
369	5..14;59..71;74..83
370	39..45;45..67;82..91
371	37..46;81..86;92..99;140..156;163..169;179..184;209..216;242..252;258..268
372	72..81;93..104;141..146
373	71..77;86..93
374	1..7;31..47;47..57;59..65
375	5..10;55..60;83..90
376	34..43;40..46;67..81
377	27..38;49..56;54..64;68..78;111..122
378	9..26;80..85;96..102
379	41..50;52..58;195..204;318..327;337..347
380	98..103
381	46..51
382	67..72
383	4..12;25..33;73..81
384	42..54;69..77;117..127;125..141
385	105..112;146..155;181..187;200..214;230..243;254..264;261..272;290..295
386	46..53;92..100;107..114;116..123;126..132;176..182
387	5..12;44..50;158..167;190..195
388	16..25;31..37;73..79;96..101;105..110;158..166;215..224;240..247;273..280;299..306;321..334;347..352;370..375
389	28..35;72..81;93..100;101..113;183..192;218..228;284..300;318..330;339..344;443..454;481..489;491..500
390	1..8;39..48
391	104..114;157..162
392	28..34;31..36;46..55;77..85
393	104..114;188..224
394	44..51;107..114
395	2..15;51..61;82..88;104..114;249..259;286..298;333..340;361..367
396	24..31;54..65;79..85;92..99;180..186;216..221
397	20..33;31..41;67..75;82..89;168..173
398	58..65;93..101;135..143;198..203
399	11..17;37..48;71..79;94..100;99..112;132..144;161..173;173..180
400	21..31;65..84;91..99
401	1..9;11..27;78..83;88..96;107..112;112..121;135..141;147..153;164..170
402	1..9;11..35;83..94;106..111;111..120;134..140;146..152;163..169
403	2..15;51..61;82..88;104..114;249..259;286..298;333..340;361..367
405	104..114;188..224
406	1..7;61..71;77..83;80..85;163..179;204..211;210..225;231..238;254..259;272..278;305..321;347..356
408	53..77;120..130;144..159;159..169;196..202;266..272;331..345
409	1..9;73..81;226..236

410	12..21;37..49;78..85
411	22..36;151..156;161..170
412	56..65;136..146;179..186;192..207;227..236
413	8..14;92..99;184..191
414	1..10;16..30;71..77;95..105;102..113;123..130;137..145;165..171;199..205;217..226;286..295;309..314;360..376;384..389
415	19..34;100..107;115..123;143..149;154..164;168..175;176..189;217..226;224..239;249..257;264..270;278..290;294..303;328..336;347..356;374..384;391..396;444..449;453..460;476..481
416	105..112;147..155;181..187;200..214;230..243;254..264;261..272;290..295
417	8..14
418	8..14;92..99;227..236;260..273;293..300
419	8..16;117..122
420	27..35;71..83;91..97;137..146
421	2..11;77..82
422	20..26;61..69
423	9..19;25..34;47..60;57..65;87..92;106..116;126..134
424	6..18;21..34;53..62;79..86;89..95;108..113;128..149
425	2..11;29..46;47..55
426	5..26;33..68
427	1..11;19..38;38..49;52..60;90..98;100..105;129..137;155..160
428	23..31;47..61;65..72;87..94
429	3..12;31..39;59..69;82..87
430	81..91
431	36..44;83..89
432	28..42;56..76;110..117
433	5..14;43..49
434	9..17;15..21;61..70;80..89
435	1..24;32..40;52..60
436	1..11;17..30;29..43
437	20..30;28..35;40..50
438	10..28;76..85;91..99;107..112
439	86..92;104..110
440	36..42;47..57;63..71;106..121
441	19..25;27..44
442	1..14;21..29;36..42;44..51;62..81;101..109;111..119;138..149
443	10..18;25..31;33..40;51..70;89..94
444	3..8;19..26;32..44
445	1..11;19..38;38..49;52..60;130..139
446	12..20;28..37;43..66;90..102
447	15..20;24..31;36..47;68..82;88..96
448	29..45;83..91;88..94;132..144
449	22..33;54..64;86..96;102..108
450	27..39;47..60;101..107;155..164;270..281;287..300;306..312;327..332
451	13..24;60..70;77..83
452	8..14;74..82
453	6..12;63..78;77..91;97..103;102..108

454	32..44;66..72;101..114;166..174;209..235;243..252;258..263
455	69..76;131..139;164..173
456	54..63;95..103;187..202;211..216;249..261
457	14..21;31..45;80..88;187..194;347..353
458	47..62;79..86
459	1..8;27..37;90..97;99..106;123..140;145..163
460	8..17;35..45;131..139;162..169;175..180
461	1..6;13..23;58..66;89..101
462	44..53;86..93
463	62..70
464	50..57;59..69;67..73;79..95
465	10..17;23..44
466	3..15;71..78;110..121;125..131;259..269;296..306;312..318;340..346;353..363;370..379;407..412;417..425;448..453;483..493
467	5..12;20..30;70..78;82..100;106..115;129..135
468	8..16;22..31;36..45;75..84
469	14..23;98..105;106..116
470	11..23;26..31;54..62;101..107
471	23..29;66..81
472	23..29;93..100
473	8..25;79..89;103..109
474	37..45;80..89;94..101;125..130
475	37..45;80..89;94..101;125..130
476	7..26;23..36;36..45;78..83;80..85
477	45..53
478	1..7;16..22;78..93;96..102
479	24..33;41..50;61..80;93..100;129..136;160..170;199..208;267..276;325..335
480	5..14;43..51;102..116
481	2..15;16..24;53..62;87..97;100..109;109..133;145..152
482	2..15;16..24;53..62;87..97;100..109;109..133;145..152;168..176

Table VIII

Seq Id No	Chromosomal location
2	16p11-p13
12	16p11-p13
22	7q35-q36
25	chr.19
34	chr.17
35	6p21.3
40	chr.20
42	12p13.3
45	12p13.3
51	12p
56	22q11.2-q13.2
57	12p13
60	chr.10
62	chr.17
65	Xq13
67	chr.14
70	chr.7(1);7q11.23-q21.1(1)
71	chr.7(1);7q11.23-q21.1(1)
73	6p21.3
74	6p21.3
87	19q13.1
88	7q21-q22
94	17q11.2
99	6q21
101	6p11.2-p21.3
103	chr.17
106	6q15-q16.3
107	16p13.3
108	12q
113	1p33-p34.3
125	6p22.1-p22.3
126	16p13.3
127	14q11.2
135	22q11.2-q13.2
138	chr.3
141	12q24.1
146	3p21.3
147	chr.2
149	chr.17
150	21q
152	21q
154	20q12-q13.11
155	11p15.5
160	19q13.2
161	19q13.2
162	20q12-q13.11
164	21q
166	21q

170	6p12.1-p21.1
172	21q
173	chr.19
176	21q
177	21q
179	chr.6
183	chr.7
185	Xq21.3-q22.3
186	chr.20
192	11q12.2
195	chr.20
196	20q13.1-q13.2
197	7p15-p21
198	19q13.3
199	chr.2
201	Xq22.1-q23
202	Xq22.1-q23
204	chr.20
205	chr.5
206	chr.2
208	chr.5
214	chr.12
220	Xq28
224	chr.7
227	chr.14
230	chr.7
231	chr.7
238	19p13.3

Table IX

Seq Id No	Tissue distribution
1	Br:28;FB:25;FK:9;Ov:17;Pl:12;Pr:4;SC:2;SG:4;Te:9
2	Br:2;CP:1;FB:5;FK:1;Pl:3;Pr:10;SG:1
3	Br:1;CP:1;FB:33;FK:13;Li:2;Ov:19;PG:12;Pl:27;Pr:15;SG:9;SI:12
4	AG:1;CP:1;LG:1;Pr:3;Te:1
5	Pa:4;Pr:2
6	Li:1;Pa:4;Pr:3
7	Br:9;Pr:1;Te:3
8	Br:4;FB:1;Pr:3;SG:8
9	Br:4;Ce:1;Co:1;DM:4;FB:33;FK:16;He:3;Ki:6;LC:2;LG:4;Li:2;Lu:2;Ly:1;Ov:36;Pa:16;Pl:2;Pr:4;SC:2;SI:1;SN:1;Sp:1;UC:3;Ut:1
10	Br:1;CP:1;Pr:4;SG:2
11	Pr:2;SG:4
12	Br:1;CP:1;FB:5;FK:1;Pl:3;Pr:9;SG:1
13	FL:4;Li:4
14	Li:4;Te:3
15	Te:1
16	Li:3;Te:6
17	Ce:1;FB:6;Li:1;Pl:5;Te:16
18	Li:7;Te:6
19	Li:27;Te:9
20	Li:1;Te:3
21	Te:3
22	Te:3
23	Li:1;Te:6
24	Li:2;Te:2
25	Te:8
26	Te:5
27	LC:1;Te:2
28	Li:1;Te:2
29	AG:2;BM:1;Br:16;CP:1;Co:2;DM:1;FB:45;FK:62;FL:1;HP:3;LC:1;Li:2;Mu:1;Ov:2;Pr:10;SI:5;SN:3;Te:9;UC:1
30	Li:2
31	Br:3;CP:1;FB:1;FK:6;Pr:1;Te:2
32	Br:1;CP:1;Ce:6;Ov:1;Te:2
33	FK:5;SC:1
34	Br:1;FB:2;FK:48;Pl:2;SN:1
35	Te:1
36	FB:5;Pr:1;SN:1
37	FB:3;FK:1;Li:1;SG:5
38	FB:10
39	FB:3
40	Br:1;DM:1;FL:1;Pl:4;SG:13
41	FB:3;FK:1;Li:1;SG:5

42	BM:1;SG:19
43	SG:1
44	CP:1;FB:1;Mu:2;Pl:9;SG:7
45	BM:1;SG:20
46	BM:1;DM:1;FB:5;FK:6;FL:1;He:1;Ki:2;Ov:9;Pl:1;SG:1;Si:1;Te:1
47	Br:4;FB:4;Pr:3;SG:8
48	Br:12;Ce:1;Co:1;FB:5;FK:4;FL:5;HP:1;Ki:1;LC:1;Li:6;Ov:8;Pl:105;SC:1;SG:8;Te:4
49	Br:7;Ce:1;Co:1;FK:4;HP:1;Ki:1;LC:1;Li:5;Ov:8;Pl:5;SC:1;Te:1
50	AG:1;CP:4;Ce:1;DM:2;FB:6;FK:4;FL:2;HP:2;LC:1;LG:3;Li:31;Lu:3;Mu:1;Ov:25;Pl:15;Pr:20;SC:1;Te:75;UC:5;Ut:1
51	FL:1
52	Br:2;CP:1;FB:3;FK:1;FL:5;LC:2;Pl:1;Pr:2;UC:2
53	Br:3;FK:4;FL:4;HP:1;Li:3;Pl:11;SG:1;Te:1
54	Br:15;Ce:1;FB:10;FK:10;FL:1;He:1;Ki:6;LC:1;Li:4;Ov:32;Pa:3;Pl:2;Pr:4;SC:1;SN:2;Sp:4;Te:8;UC:1;Ut:1
55	FL:2
56	Br:1;FB:1;FL:1;Te:1
57	FL:4
58	FL:1;Li:1
59	FB:3;FK:1;Li:1;SG:5
60	Br:1;FB:1;FL:1;Pr:2
61	Br:2;Pl:1
62	Br:6;CP:1;Ce:7;FB:37;FK:4;FL:1;Pl:6;Pr:1;SG:3;SN:3;Te:1;UC:1
63	Br:10
64	Br:2;CP:2
65	Br:1;FB:11;LG:1;Th:1
66	Br:30;Ce:1;Co:1;FB:60;FK:15;FL:3;HP:1;Ki:1;LC:1;Li:6;Ov:57;PG:9;Pl:145;Pr:21;SC:1;Si:4;Te:4
67	Br:4;CP:1;FB:14;Ki:1;Li:1;Lu:2;Pr:1;Te:1
68	Br:10
69	AG:1;Br:48;FB:3;FK:5;HP:1;He:1;Li:1;Pl:11;SC:2;SG:1;Te:2;Ut:1
70	Br:11;DM:1;He:1
71	DM:1;He:1
72	Br:9;Pr:1;Te:2
73	Br:8;Pr:1
74	Br:5
76	AG:1;Br:49;FB:4;FK:5;HP:1;He:1;LC:1;Li:1;Pl:11;SC:2;SG:1;Te:2;Ut:1
77	Br:2;FK:2;HP:1;LC:1;Li:2;Ov:14;Pl:1;Pr:14;Te:5
78	Br:9;Ce:1;DM:1;FB:21;FK:18;FL:1;HP:1;He:1;Ki:9;LC:2;LG:4;Li:2;Lu:2;Ov:34;Pl:3;Pr:4;SC:1;Si:2;SN:2;Sp:1;Ut:1
79	Pr:1
80	Br:9;CP:2;Co:1;DM:6;FB:1;FK:6;He:2;Ki:4;LC:1;LG:1;Li:7;Ov:40;Pa:1;Pl:2;Pr:1;SN:2;Sp:1;Te:12;UC:1;Ut:3
81	FK:1;Te:1
82	Li:1

83	Br:2;CP:1;FB:10;FK:2;Ki:3;Li:7;Ov:10;SC:1;SN:1;Te:1;UC:1
84	Br:5;FB:14;FK:9;Li:6;Ov:17;SG:8;Te:8
85	Li:6;Te:2
86	Li:2;Te:2
87	Br:1;FB:35;FK:31;Li:20;Ov:37;PG:5;Pl:69;Sl:5;Te:5
88	Li:1;Pr:1;Te:7;Ut:2
89	Te:1
90	Te:2
91	FB:15;FK:3;Li:2;Ov:17;Pr:4;SG:7;Te:4
92	Te:2
93	Br:4;FB:1;SN:1;Te:2
94	Te:1
95	Li:2
96	AG:1;Br:1;FB:1
97	FK:5;Te:2
98	Te:3
99	Br:3;FB:29;FK:1;Li:10;Ov:1;Pl:16;Pr:2;SG:1;Te:49
100	Br:2;FB:3;FK:1;Ov:3;Te:1
101	Br:10;FB:34;FK:1;Ov:1;Pl:85;Pr:1;Ut:1
102	FB:6
103	FB:6;Li:3;Pl:1;Pr:1;SG:1;Te:7
104	Br:26;CP:1;FB:8;FK:131;Pl:20;Pr:20
105	Br:3;CP:2;DM:2;FB:11;FK:3;LG:2;Ov:1;Pl:6;SC:2;SG:1;SN:4
106	FB:4
107	Br:3;FB:50;FK:59;FL:3;Pr:1
108	Br:1;FB:8;Li:1;Lu:1
109	FB:11;Pr:21
110	Br:14;Ce:1;FB:5;FK:5;FL:1;HP:1;He:2;Ki:3;LC:1;Li:4;Lu:1;Ov:7;Pl:2;Pr:1;Sl:1;Sp:1;UC:1
111	Br:1;Ce:1;FK:1;HP:1;He:2;Ki:3;LC:1;Li:3;Lu:1;Ov:7;Sp:1;UC:1
112	Br:1;HP:1;Lu:1;Pr:1;SG:9;Ut:1
113	HP:1;SG:4
114	FK:9
115	AG:1;Br:3;CP:1;FB:14;FK:19;FL:1;HP:1;Pr:1;SG:1
116	Br:5;CP:1;Ce:1;Co:1;DM:5;FK:3;FL:1;LC:3;LG:1;Lu:1;Ov:23;Pl:1;Te:8;UC:2;Ut:4
117	Br:1;Ce:1;FB:1;FK:1;FL:2;Pl:3;SN:1;Te:1;UC:1
118	CP:1;DM:1;FB:5;FK:2;FL:2;He:1;Lu:1;Ly:1;Ov:23;Pr:1;SN:2;Sp:2;Ut:1
119	Li:2;Te:7
120	Br:6;Co:2;FB:1;FK:6;FL:1;Ov:3;Pl:32;Pr:1;SN:1
121	AG:1;Br:4
122	Br:5
123	Br:1
124	Br:2;Ki:2;Li:1;Ov:7;UC:1
125	Br:2;FB:1;FL:6;He:1;Li:1;Ov:1;Pl:2;Pr:10;Te:1;Th:1
126	Br:1

127	BM:1;Br:2;CP:2;FB:1;FK:3;HP:1;He:1;LG:1;Pl:1;Pr:1;SC:2;SG:2;Te:5;Ut:3
128	Br:1
129	Br:2;FB:6;Li:1;SG:3;Te:2
130	Br:25;FB:3;FL:2
131	Br:1
132	Br:1
133	Br:1
134	Br:2;SN:1
135	Br:1
136	AG:1;Br:1;FL:1
137	Br:1;Ce:1;FB:1;FK:1;FL:2;Pl:3;SN:1;Te:1;UC:1
138	Br:43
139	Br:11;CP:2;Co:1;DM:6;FB:1;FK:6;He:2;Ki:4;LC:1;LG:1;Ov:40;Pa:1;Pl:2;Pr:1;SN:2;Sp:1;Te:9;UC:1;Ut:3
140	Br:23;Ce:1;DM:3;FB:38;FK:17;FL:2;HP:1;He:1;Ki:8;LC:3;LG:2;Li:6;Lu:1;Ly:1;Ov:40;Pr:4;SC:2;SN:4;Sp:1;Te:5;UC:1;Ut:1
141	Br:39;FB:3;SN:2
142	Br:10;SN:2
143	Br:26;FK:2;HP:1;LC:1;Li:2;Ov:14;Pl:3;Pr:3;Te:5
144	Br:14;Pr:2
145	FB:12;LG:1;Pr:4;Te:1;Ut:2
146	Li:1;Ov:2;Pr:5;SG:11
147	Li:1;Te:1
148	Br:1;FB:1;Li:1;Te:1
149	Br:3;FB:5;FK:5;Li:1;Pl:8;Te:5
150	FK:6;Pr:2;SG:8
151	FK:9
152	FK:6;Pr:2;SG:9
153	Te:1
154	FB:28;Ov:4
155	Br:21;Ce:1;FB:32;FK:4
156	Br:5;CP:1;FB:16;FK:3;He:1;Ki:5;Li:1;Ov:15;Pl:3;SG:2;SI:1;Sp:1;UC:1
157	FB:14;FK:1;FL:1;SG:1
158	FB:7
159	FB:10
160	Ce:2;FB:12
161	Ce:2
162	FB:28;Ov:2
163	FB:14;FK:1;FL:1;SG:1
164	FK:4;Pr:1;SG:9
165	Br:4;Co:1;Ki:1;Ov:2;Pr:4;SG:1
166	FK:6;Pr:2;SG:9
167	Br:1;FB:1;SG:5
168	Br:1;FB:5;FK:7;SG:1;UC:1
169	FK:2
170	FL:12

171	Br:2;FB:1;FK:1;Pl:7
172	Br:106;FB:2;Pl:7
173	Br:14;FB:1;Pl:2;Te:1
174	Br:17;He:1;Pl:1;SC:2;Te:1
175	Br:14;Pr:2
176	Br:106;FB:2;Pl:7
177	Br:114;FB:7;FK:7;Ov:2;Pl:7;Pr:2;Te:9
178	Br:16;CP:2;FB:2;FK:2;FL:1;Li:1;Pl:13;Pr:3;SC:1;Ut:1
179	FL:1;HP:2;Pr:2;Te:1
180	Pr:2
181	FB:1;Ov:2;Pr:1;UC:1
182	BM:1;Br:4;DM:1;FB:6;FK:6;Ki:5;LC:2;LG:1;Li:1;Lu:1;Ov:15;Pl:1;Pr:2;SC:1;Sp:2;Te:2;Ut:1
183	Br:8;CP:1;Co:2;DM:4;FB:1;FK:1;Ki:4;LC:1;Li:3;Ov:33;Pl:1;Pr:5;SC:2;SN:1;Sp:1;Te:5;UC:1;Ut:2
184	Pr:1
185	FB:2;Li:1;Ov:1;SG:7;Te:5
186	Te:3
187	Te:1
188	Br:18;CP:1;DM:5;FB:40;FK:23;FL:2;He:3;Ki:10;LC:2;LG:1;Li:13;Lu:3;Ly:2;Mu:1;Ov:54;Pl:5;Pr:14;SC:2;SG:2;SI:2;SN:4;Sp:3;Te:4;UC:4
189	Li:1;Te:1
190	Br:7;CP:1;FB:1;FK:4;FL:5;He:1;Li:1;Ov:1;Pl:2;Pr:4;SG:1
191	Li:2;Te:4
192	AG:1;Br:2;CP:1;FB:32;FK:1;Li:1;Ov:36;Pl:49;Pr:3;SC:1;SG:4;SN:4;Te:9;UC:1;Ut:2
193	FB:31;FK:75;FL:7;Ov:12;Pl:23;Pr:8;SG:3;Te:16
194	Te:2
195	Te:7
196	Te:2
197	Te:3
198	Li:10;Te:43
199	Br:35;CP:3;FB:39;FK:56;FL:7;HP:1;LG:1;Li:1;Ly:1;Ov:2;Pl:10;Pr:8;SG:1;Te:4;Ut:2
200	FB:17;FK:9;FL:5;Ov:21;Pl:41;Te:3
201	FK:16;SI:1
202	Br:1;Co:1;FB:111;FK:25;He:1;Li:4;Ov:3;Pr:6;Te:1
204	Te:7
205	Li:7;Te:28
206	FB:28;Li:2;Ov:23;PG:11;Pl:45;SG:17;SI:11;Te:9
207	FB:16;FK:1;Ov:1;SC:1;Te:1
208	FB:5
209	FB:6
210	Br:1;FB:22
211	Br:2;Ce:3;FB:6;FK:1
212	Br:1;Co:2;FB:22;FK:2;LG:2;Mu:2;Pl:2;SG:4

213	Br:2;DM:1;FB:8;FK:8;FL:1;Ki:1;LG:3;Ov:5;Pa:1;Pl:4;Pr:1;SN:2;UC:1
214	FB:7
215	FB:4
216	Ov:3;SG:3
217	Br:4;CP:2;DM:1;FB:9;FK:3;Ki:2;LC:1;LG:1;Lu:3;Ly:1;Ov:14;Pl:1;Pr:1;SC:1;SG:2;Sp:1;Te:1;Ut:1
218	FB:4;FK:2;Pl:1;Pr:1;SG:1
219	Br:7;CP:3;FB:2;FL:1;HP:4;Lu:1;Ly:2;Mu:1;Ov:3;Pl:1;Pr:1;SN:2;Te:1
220	Br:1;FL:1;Pl:2
221	Co:1;FB:2;FL:1;Li:1;Pl:2
222	FL:1;SG:2
223	Li:1;Te:1
225	Li:10
226	Li:1;Te:4
227	Li:1
228	Br:1
229	Br:3
230	Br:5;Ce:1;Co:1;DM:3;FB:1;FK:1;He:1;LC:1;LG:2;Ov:16;Pl:3;Pr:1;Te:2;Ut:1
231	Br:3;Ce:1;Co:1;DM:3;FB:1;FK:1;He:1;LC:1;LG:2;Ov:16;Pl:3;Pr:1;Te:2;Ut:1
232	AG:1;Br:17;CP:2;DM:1;FB:51;FK:9;FL:3;Li:3;Ov:3;Pl:2;Pr:10;SC:1;SG:5;Te:2;Ut:1
233	Br:13
234	Br:5
235	Br:1;Pl:1
236	Br:9
237	Br:22;DM:2;FB:17;FK:9;Ki:4;LG:1;Li:1;Lu:2;Ov:24;Pr:3;SC:1;SI:1;SN:2;Te:2
238	Br:17
239	Br:11
240	Br:28;Ce:1;DM:5;FB:52;FK:40;FL:2;HP:1;He:2;Ki:3;LC:1;LG:3;Li:1;Ly:1;Ov:28;Pl:1;Pr:5;SC:1;SI:1;SN:3;Sp:6;Te:1;UC:1;Ut:1
241	Br:4;Ce:1;DM:5;FB:5;FK:7;HP:1;He:2;Ki:3;LC:1;LG:3;Li:1;Ly:1;Ov:28;Pl:1;SC:1;SN:3;Sp:6;Te:1;UC:1;Ut:1

Table X

Seq Id No	Low frequency expression	High frequency expression
1	-	Br,Ov
2	-	Pr
3	Br,Te	Ov,PG,Pl,SI
4	-	AG
5	-	Pa
6	-	Pa
7	-	Br
8	-	SG
9	Br,Te	DM,He,Ki,Ov,Pa
10	-	Pr
11	-	SG
12	-	Pr
13	-	FL,Li
14	-	Li,Te
15	-	Te
16	-	Li,Te
17	-	Te
18	-	Li,Te
19	-	Li,Te
20	-	Te
21	-	Te
22	-	Te
23	-	Te
24	-	Li
25	-	Te
26	-	Te
27	-	LC,Te
28	-	Te
29	Pl	FK
30	-	Li
31	-	FK
32	-	Ce
33	-	FK,SC
34	FB	FK
35	-	Te
36	-	SN
37	-	SG
38	-	FB
40	-	SG
41	-	SG
42	-	BM,SG
43	-	SG

44	-	Mu,Pl,SG
45	-	BM,SG
46	-	BM,Ki,Ov
47	-	SG
48	FB,FK,Pr	Pl
49	-	Ki,Ov
50	Br,FB,FK,SG	Li,Ov,Te
51	-	FL
52	-	FL,LC,UC
53	-	Pl
54	-	Ki,Ov,Pa,Sp
55	-	FL
57	-	FL
58	-	FL
59	-	SG
62	-	Ce,FB
63	-	Br
64	-	CP
65	-	FB,Th
66	FK,SG,Te	Ov,PG,Pl
67	-	FB,Ki,Lu
68	-	Br
69	FB	Br
70	-	Br,DM,He
71	-	DM,He
72	-	Br
73	-	Br
74	-	Br
75	-	Br
76	FB	Br
77	FB	Ov,Pr
78	-	Ki,Ov
80	FB	DM,Ki,Ov
82	-	Li
83	-	Ki,Li,Ov
84	-	Ov
85	-	Li
86	-	Li
87	Br,Pr,SG	Ov,PG,Pl
88	-	Te,Ut
89	-	Te
90	-	Te
91	-	Ov
92	-	Te
93	-	SN
94	-	Te

95	-	Li
96	-	AG
97	-	FK
98	-	Te
99	FK	Te
100	-	Ov
101	FK	Pl
102	-	FB
103	-	Te
104	FB,Li,SG,Te	FK
105	-	DM,SN
106	-	FB
107	Br,Pl	FB,FK
108	-	FB,Lu
109	-	Pr
110	-	He,Ki,Ov
111	-	Ce,He,Ki,Lu,Ov
112	-	Lu,SG
113	-	HP,SG
114	-	FK
115	-	FK
116	FB	DM,LC,Ov,Ut
117	-	Ce,UC
118	-	Ov,Sp
119	-	Te
120	FB	Co,Pl
121	-	AG,Br
122	-	Br
124	-	Ki,Ov
125	-	FL,Pr,Th
127	-	BM,SC,Ut
130	-	Br
134	-	SN
136	-	AG
137	-	Ce,UC
138	FB	Br
139	FB	DM,Ki,Ov,Ut
140	Pl	Ki,Ov
141	-	Br
142	-	Br,SN
143	FB	Br,Ov
144	-	Br
145	-	FB,Ut
146	-	SG
149	-	Pl
150	-	FK,SG

151	-	FK
152	-	FK,SG
153	-	Te
154	-	FB,Ov
155	-	Br,FB
156	-	Ki,Ov
157	-	FB
158	-	FB
159	-	FB
160	-	Ce,FB
161	-	Ce
162	-	FB
163	-	FB
164	-	SG
165	-	Co,Ki,Ov
166	-	FK,SG
167	-	SG
168	-	FK
169	-	FK
170	-	FL
171	-	Pl
172	FB,FK,Pr	Br
173	-	Br
174	-	Br,He,SC
175	-	Br
176	FB,FK,Pr	Br
177	FB	Br
178	-	Br,Pl
179	-	HP
180	-	Pr
181	-	Ov,UC
182	-	Ki,Ov,Sp
183	FB	DM,Ki,Ov
185	-	SG,Te
186	-	Te
187	-	Te
188	Pl	DM,Ki,Ov
190	-	FL
191	-	Te
192	Br,FK	Ov,Pl
193	Br	FK,Ov
194	-	Te
195	-	Te
196	-	Te
197	-	Te
198	FB	Li,Te

199	-	FK
200	Br	Ov,Pl
201	-	FK
202	-	FK
203	Br,Pl	FB
204	-	Te
205	-	Li,Te
206	Br,FK,Pr	Ov,PG,Pl,SG,SI
207	-	FB
208	-	FB
209	-	FB
210	-	FB
211	-	Ce
212	-	Co,FB,Mu
213	-	Ki,LG,Ov
214	-	FB
215	-	FB
216	-	Ov,SG
217	-	Ki,Lu,Ov
218	-	Pr
219	-	CP,HP,Ly,Ov,SN
221	-	Co
222	-	SG
223	-	SG
225	-	Li
226	-	Te
227	-	Li
229	-	Br
230	-	DM,Ov
231	-	DM,Ov
232	-	FB
233	-	Br
234	-	Br
236	-	Br
237	-	Ki,Lu,Ov
238	-	Br
239	-	Br
240	Pl,Te	DM,FK,Ki,Ov,Sp
241	FB	DM,He,Ki,Ov,Sp

Table XI

Seq Id No	Subcellular localization
7	nuclear
13	extracellular, including cell wall
20	mitochondrial
21	nuclear
26	nuclear
35	nuclear
37	endoplasmic reticulum
38	extracellular, including cell wall
39	endoplasmic reticulum
41	endoplasmic reticulum
59	endoplasmic reticulum
70	nuclear
71	nuclear
72	nuclear
78	nuclear
98	nuclear
99	nuclear
105	mitochondrial
108	endoplasmic reticulum
116	mitochondrial
117	mitochondrial
134	nuclear
135	nuclear
137	mitochondrial
159	nuclear
160	nuclear
161	nuclear
171	nuclear
178	endoplasmic reticulum
182	nuclear
184	nuclear
185	endoplasmic reticulum
186	nuclear
187	nuclear
188	nuclear
194	nuclear
195	nuclear
196	nuclear
200	mitochondrial
204	nuclear
205	nuclear
206	nuclear

211	nuclear
212	nuclear
213	nuclear
214	endoplasmic reticulum
215	endoplasmic reticulum
216	endoplasmic reticulum
218	nuclear
220	endoplasmic reticulum
224	nuclear
225	nuclear
230	mitochondrial
231	mitochondrial
238	cytoplasmic

Table XII

Seq Id No in priority applications	Internal designati n	Seq Id N in present application
119	119-003-4-0-C2-CS	1
220	105-016-1-0-D3-CS	2
345	105-016-3-0-G10-CS	3
334	105-026-1-0-A5-CS	4
159	105-031-1-0-A11-CS	5
219	105-031-2-0-D3-CS	6
250	105-035-2-0-C6-CS	7
217	105-037-2-0-H11-CS	8
340	105-053-4-0-E8-CS	9
115	105-074-3-0-H10-CS	10
31	105-089-3-0-G10-CS	11
198	105-095-2-0-G11-CS	12
154	106-006-1-0-E3-CS	13
366	106-037-1-0-E9-CS.cor	14
366	106-037-1-0-E9-CS.fr	15
79	106-043-4-0-H3-CS	16
95	110-007-1-0-C7-CS	17
364	114-016-1-0-H8-CS	18
246	116-004-3-0-A6-CS	19
187	116-054-3-0-E6-CS	20
203	116-055-1-0-A3-CS	21
298	116-055-2-0-F7-CS	22
277	116-088-4-0-A9-CS	23
41	116-091-1-0-D9-CS	24
353	116-110-2-0-F4-CS	25
78	116-111-1-0-H9-CS	26
245	116-111-4-0-B3-CS	27
104	116-115-2-0-F8-CS	28
259	116-119-3-0-H5-CS	29
269	117-001-5-0-G3-CS	30
166	145-25-3-0-B4-CS.cor	31
166	145-25-3-0-B4-CS.fr	32
169	145-56-3-0-D5-CS	33
312	145-59-2-0-A7-CS	34
273	157-15-4-0-B11-CS	35
190	160-103-1-0-F11-CS	36
244	160-37-2-0-H7-CS	37
151	160-58-3-0-H3-CS	38
149	160-75-4-0-A9-CS	39
307	174-10-2-0-F8-CS	40
264	174-33-3-0-F6-CS	41

168	174-38-1-0-B6-CS	42
202	174-38-3-0-C9-CS	43
28	174-39-2-0-A3-CS	44
331	174-41-1-0-A6-CS	45
258	174-5-3-0-H7-CS	46
84	174-7-4-0-H1-CS	47
294	175-1-3-0-E5-CS.cor	48
294	175-1-3-0-E5-CS.fr	49
310	180-19-4-0-F4-CS	50
311	181-10-1-0-D10-CS	51
263	181-16-1-0-G7-CS	52
304	181-16-2-0-A7-CS	53
109	181-20-3-0-B5-CS	54
121	181-3-3-0-B8-CS	55
181	181-3-3-0-C9-CS	56
191	182-1-2-0-D12-CS	57
193	184-1-4-0-C11-CS	58
192	184-4-1-0-A11-CS	59
116	187-12-4-0-A8-CS	60
268	187-2-2-0-A3-CS	61
123	187-31-0-0-f12-CS	62
234	187-34-0-0-l12-CS	63
185	187-37-0-0-c10-CS	64
279	187-38-0-0-l10-CS	65
114	187-39-0-0-k12-CS	66
211	187-41-0-0-i21-CS	67
236	188-11-1-0-B3-CS	68
35	188-18-4-0-A9-CS	69
299	188-28-4-0-B12-CS.cor	70
299	188-28-4-0-B12-CS.fr	71
72	188-28-4-0-D4-CS	72
242	188-41-1-0-B8-CS.cor	73
242	188-41-1-0-B8-CS.fr	74
173	188-45-1-0-D9-CS	75
106	188-9-2-0-E1-CS	76
130	105-079-3-0-A11-CS	77
323	105-092-1-0-H7-CS	78
160	105-141-4-0-H9-CS	79
272	109-013-1-0-B9-CS	80
226	110-008-4-0-D9-CS	81
333	114-001-3-0-A2-CS	82
315	114-028-2-0-C1-CS	83
300	114-032-1-0-H10-CS	84
57	114-043-2-0-A10-CS	85
137	114-044-1-0-C5-CS	86
107	116-003-3-0-D10-CS	87

164	116-003-3-0-G12-CS	88
108	116-011-2-0-F11-CS	89
101	116-033-3-0-E4-CS	90
157	116-041-4-0-B6-CS	91
75	116-044-2-0-C4-CS	92
322	116-075-1-0-E6-CS	93
124	116-094-4-0-G5-CS	94
289	117-005-3-0-F2-CS	95
122	121-007-3-0-D9-CS	96
208	145-91-3-0-D10-CS	97
282	157-17-1-0-F4-CS	98
129	160-11-3-0-G8-CS	99
317	160-24-1-0-F12-CS	100
308	160-24-2-0-E9-CS	101
25	160-25-4-0-D2-CS	102
243	160-31-3-0-A11-CS	103
346	160-32-1-0-F6-CS	104
60	160-37-1-0-A3-CS	105
305	160-40-3-0-E9-CS	106
48	160-58-3-0-E4-CS	107
238	160-85-3-0-D4-CS	108
251	160-95-3-0-A11-CS	109
196	162-10-4-0-F9-CS.cor	110
196	162-10-4-0-F9-CS.fr	111
347	174-13-2-0-E4-CS	112
77	174-46-2-0-B11-CS	113
188	179-8-2-0-A6-CS	114
235	180-22-3-0-B6-CS	115
45	181-13-1-0-F7-CS	116
265	181-15-4-0-F7-CS	117
280	181-20-1-0-G7-CS	118
281	184-15-3-0-D1-CS	119
39	187-12-2-0-G11-CS	120
165	187-2-2-0-A12-CS	121
326	187-30-0-0-k23-CS	122
330	187-36-0-0-e19-CS	123
368	187-38-0-0-d22-CS	124
71	187-39-0-0-b9-CS	125
224	187-39-0-0-g6-CS	126
90	187-45-0-0-l18-CS	127
216	187-45-0-0-m21-CS	128
83	187-45-0-0-n8-CS	129
342	187-46-0-0-f23-CS	130
262	187-5-1-0-A12-CS	131
257	187-5-1-0-F6-CS	132
293	187-5-2-0-B2-CS	133

231	187-5-3-0-D5-CS	134
287	187-51-0-0-f9-CS	135
325	187-6-1-0-B9-CS	136
309	187-6-4-0-C10-CS	137
359	188-19-2-0-C8-CS	138
68	188-22-4-0-G6-CS	139
233	188-28-4-0-D11-CS	140
369	188-29-1-0-E10-CS	141
155	188-34-4-0-E5-CS	142
327	188-9-3-0-A5-CS	143
283	105-021-3-0-C3-CS	144
29	105-037-4-0-H12-CS	145
100	105-073-2-0-A7-CS	146
99	109-002-4-0-C6-CS	147
360	109-003-1-0-G4-CS	148
321	116-118-4-0-A8-CS	149
120	145-52-2-0-D12-CS	150
230	145-7-2-0-G5-CS	151
177	145-7-3-0-D3-CS	152
43	157-17-2-0-C1-CS	153
352	160-101-3-0-H2-CS	154
47	160-12-1-0-D10-CS	155
195	160-28-4-0-C4-CS	156
344	160-31-3-0-E4-CS	157
61	160-40-1-0-H4-CS	158
237	160-54-1-0-F7-CS	159
32	160-88-3-0-A8-CS.cor	160
32	160-88-3-0-A8-CS.fr	161
97	160-99-4-0-E4-CS	162
249	161-5-4-0-B6-CS	163
218	174-17-1-0-D6-CS	164
266	174-32-4-0-F8-CS	165
161	174-38-4-0-D11-CS	166
113	174-8-2-0-C10-CS	167
255	179-14-2-0-F11-CS	168
24	179-9-4-0-B8-CS	169
128	181-10-1-0-C9-CS	170
58	187-5-3-0-C7-CS	171
358	188-26-4-0-F5-CS	172
171	188-27-3-0-G1-CS	173
98	188-29-2-0-H1-CS	174
133	188-31-1-0-E6-CS	175
49	188-45-1-0-D3-CS	176
42	188-5-1-0-H6-CS	177
148	188-9-1-0-C10-CS	178
319	105-016-3-0-C5-CS	179

254	105-026-4-0-D9-CS	180
55	105-053-2-0-D9-CS	181
89	105-069-3-0-A11-CS	182
212	105-076-4-0-F6-CS	183
143	105-135-2-0-F9-CS	184
179	106-023-4-0-F6-CS	185
54	110-001-3-0-C11-CS	186
354	110-002-3-0-F9-CS	187
297	114-019-3-0-D9-CS	188
291	114-029-1-0-C6-CS	189
30	114-032-4-0-B1-CS	190
363	114-070-2-0-H4-CS	191
194	116-016-3-0-F11-CS	192
76	116-022-4-0-G2-CS	193
232	116-052-2-0-H8-CS	194
139	116-053-4-0-B4-CS	195
145	116-094-3-0-H2-CS	196
206	116-112-4-0-C7-CS	197
88	116-123-3-0-F12-CS	198
69	123-008-1-0-C5-CS	199
285	145-53-2-0-H8-CS	200
96	145-57-2-0-C9-CS.cor	201
96	145-57-2-0-C9-CS.fr	202
26	145-7-3-0-B12-CS	203
261	157-12-2-0-D1-CS	204
349	157-16-2-0-D5-CS	205
92	157-18-2-0-A7-CS	206
81	160-103-1-0-B10-CS	207
74	160-104-4-0-F3-CS	208
66	160-22-2-0-D10-CS	209
270	160-24-3-0-F12-CS	210
318	160-3-2-0-H3-CS	211
215	160-58-2-0-A2-CS	212
214	160-73-1-0-B4-CS	213
94	160-75-4-0-E6-CS	214
73	160-97-3-0-E9-CS	215
296	174-1-4-0-E9-CS	216
127	174-12-4-0-C2-CS	217
350	180-19-4-0-H2-CS	218
320	181-10-4-0-G12-CS	219
292	181-3-2-0-F6-CS	220
125	181-4-4-0-A12-CS	221
91	181-9-2-0-F12-CS.cor	222
91	181-9-2-0-F12-CS.fr	223
338	184-13-3-0-E11-CS	224
102	184-4-2-0-D3-CS	225

105	184-7-1-0-E7-CS	226
112	184-8-4-0-G9-CS	227
267	187-10-3-0-G9-CS	228
260	187-32-0-0-m20-CS	229
343	187-32-0-0-n21-CS.cor	230
343	187-32-0-0-n21-CS.fr	231
341	187-4-2-0-E6-CS	232
59	187-40-0-0-i15-CS	233
111	187-47-0-0-g24-CS	234
40	187-9-3-0-A2-CS	235
355	188-26-4-0-H1-CS	236
37	188-35-3-0-G9-CS	237
180	188-38-4-0-D8-CS	238
152	188-41-1-0-E6-CS	239
93	188-42-2-0-F3-CS.cor	240
93	188-42-2-0-F3-CS.fr	241

WHAT IS CLAIMED IS:

1. An isolated polynucleotide, said polynucleotide comprising a nucleic acid sequence encoding:
 - 5 i) a polypeptide comprising an amino acid sequence having at least about 80% identity to any one of the sequences shown as SEQ ID NOs:242-482 or any one of the sequences of polypeptides encoded by the clone inserts of the deposited clone pool; or
 - 10 ii) a biologically active fragment of said polypeptide.
2. The polynucleotide of claim 1, wherein said polypeptide comprises any one of the sequences shown as SEQ ID NOs:242-482 or any one of the sequences of the polypeptides encoded by the clone inserts of the deposited clone pool.
- 15 3. The polynucleotide of claim 1, wherein said polypeptide comprises a signal peptide.
4. The polynucleotide of claim 1, wherein said polypeptide is a mature protein.
5. The polynucleotide of claim 1, wherein said nucleic acid sequence has at least about
20 80% identity over at least about 100 contiguous nucleotides to any one of the sequences shown as SEQ ID NOs:1-241 or any one of the sequences of the clone inserts of the deposited clone pool.
6. The polynucleotide of claim 1, wherein said polynucleotide hybridizes under stringent conditions to a polynucleotide comprising any one of the sequences shown as SEQ ID
25 NOs:1-241 or any one of the sequences of the clone inserts of the deposited clone pool.
7. The polynucleotide of claim 5, wherein said nucleic acid sequence comprises any one of the sequences shown as SEQ ID NOs:1-241 or any one the sequences of the clone inserts of the deposited clone pool.
30
8. The polynucleotide of claim 1, wherein said polynucleotide is operably linked to a promoter.
9. An expression vector comprising the polynucleotide of claim 8.
- 35 10. A host cell recombinant for the polynucleotide of claim 1.

11. A non-human transgenic animal comprising the host cell of claim 10.
12. A method of making a GENSET polypeptide, said method comprising
- 5 a) providing a population of host cells comprising the polynucleotide of claim 8; and
- b) culturing said population of host cells under conditions conducive to the production of said polypeptide within said host cells.
13. The method of claim 12, further comprising purifying said polypeptide from said
- 10 population of host cells.
14. A method of making a GENSET polypeptide, said method comprising
- a) providing a population of cells comprising the polynucleotide of claim 8;
- 15 b) culturing said population of cells under conditions conducive to the production of said polypeptide within said cells; and
- c) purifying said polypeptide from said population of cells.
15. An isolated polynucleotide, said polynucleotide comprising a nucleic acid sequence
- 20 having at least about 80% identity over at least about 100 contiguous nucleotides to any one of the sequences shown as SEQ ID NOs:1-241 or any one of the sequences of the clone inserts of the deposited clone pool.
16. The polynucleotide of claim 15, wherein said polynucleotide hybridizes under
- 25 stringent conditions to a polynucleotide comprising any one of the sequences shown as SEQ ID NOs:1-241 or any one of the sequences of the clone inserts of the deposited clone pool.
17. The polynucleotide of claim 15, wherein said polynucleotide comprises any one of the sequences shown as SEQ ID NOs:1-241 or any one of the sequences of the clone inserts of the
- 30 deposited clone pool.
18. A biologically active polypeptide encoded by the polynucleotide of claim 15.
19. An isolated polypeptide or biologically active fragment thereof, said polypeptide
- 35 comprising an amino acid sequence having at least about 80% sequence identity to any one of the sequences shown as SEQ ID NOs:242-482 or any one of the sequences of polypeptides encoded by the clone inserts of the deposited clone pool.

20. The polypeptide of claim 19, wherein said polypeptide is selectively recognized by an antibody raised against an antigenic polypeptide, or an antigenic fragment thereof, said antigenic polypeptide comprising any one of the sequences shown as SEQ ID NOs:242-482 or any one of the sequences of polypeptides encoded by the clone inserts of the deposited clone pool.

21. The polypeptide of claim 19, wherein said polypeptide comprises any one of the sequences shown as SEQ ID NOs:242-482 or any one of the sequences of polypeptides encoded by the clone inserts of the deposited clone pool.

22. The polypeptide of claim 19, wherein said polypeptide comprises a signal peptide.

23. The polypeptide of claim 19, wherein said polypeptide is a mature protein.

24. An antibody that specifically binds to the polypeptide of claim 19.

25. A method of determining whether a GENSET gene is expressed within a mammal, said method comprising the steps of:

a) providing a biological sample from said mammal

b) contacting said biological sample with either of:

i) a polynucleotide that hybridizes under stringent conditions to the polynucleotide of claim 1; or

ii) a polypeptide that specifically binds to the polypeptide of claim 19; and

c) detecting the presence or absence of hybridization between said polynucleotide and an RNA species within said sample, or the presence or absence of binding of said polypeptide to a protein within said sample;

wherein a detection of said hybridization or of said binding indicates that said GENSET gene is expressed within said mammal.

26. The method of claim 25, wherein said polynucleotide is a primer, and wherein said hybridization is detected by detecting the presence of an amplification product comprising the sequence of said primer.

27. The method of claim 25, wherein said polypeptide is an antibody.

28. A method of determining whether a mammal has an elevated or reduced level of GENSET gene expression, said method comprising the steps of :

- a) providing a biological sample from said mammal; and
- b) comparing the amount of the polypeptide of claim 19, or of an RNA species encoding said polypeptide, within said biological sample with a level detected in or expected from a control sample;

5 wherein an increased amount of said polypeptide or said RNA species within said biological sample compared to said level detected in or expected from said control sample indicates that said mammal has an elevated level of said GENSET gene expression, and wherein a decreased amount of said polypeptide or said RNA species within said biological sample compared to said level detected in or expected from said control sample indicates that said mammal has a reduced level of
10 said GENSET gene expression.

29. A method of identifying a candidate modulator of a GENSET polypeptide, said method comprising :

- a) contacting the polypeptide of claim 18 with a test compound; and
- 15 b) determining whether said compound specifically binds to said
polypeptide;

 wherein a detection that said compound specifically binds to said polypeptide indicates that said compound is a candidate modulator of said GENSET polypeptide.

20

1/5

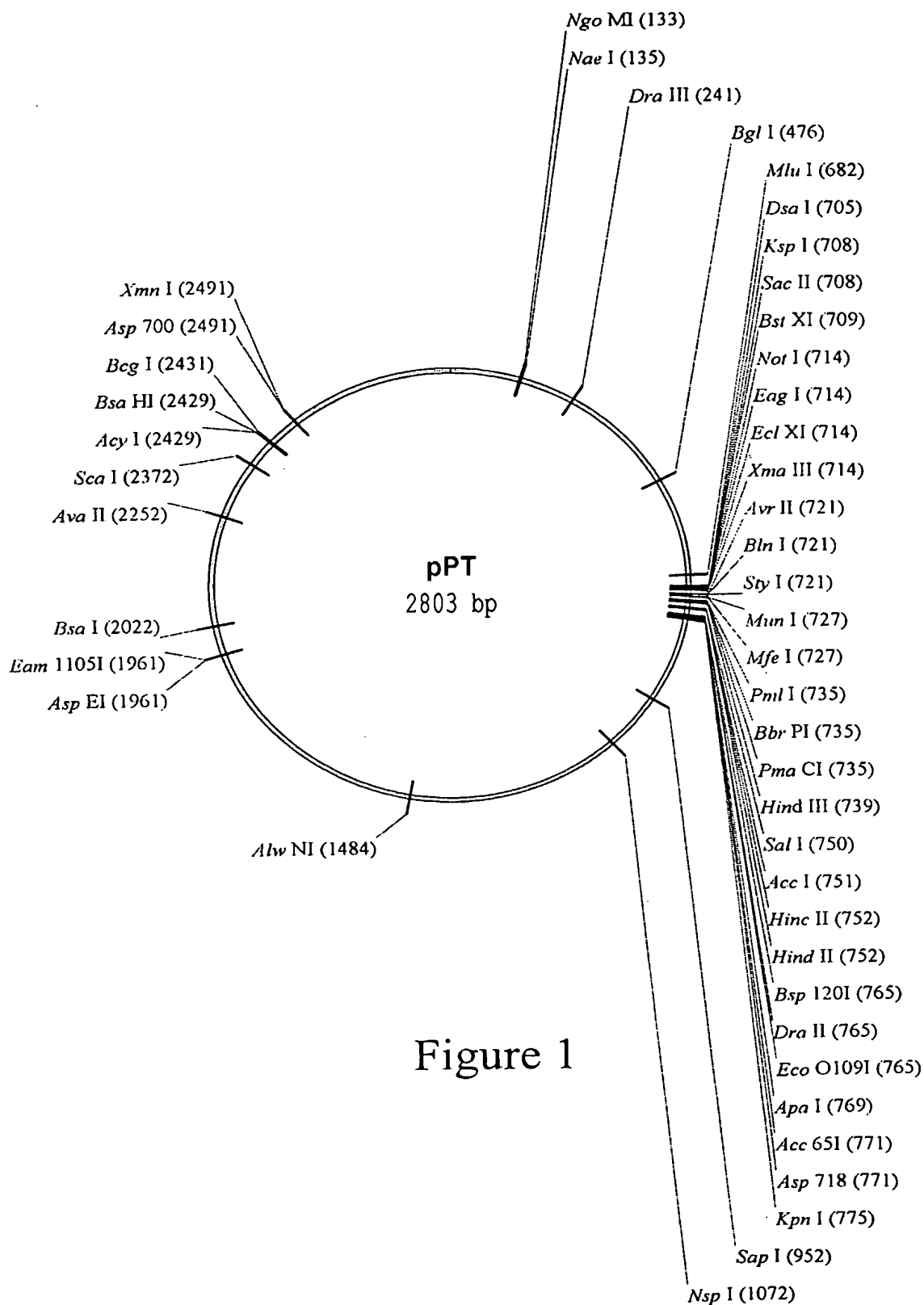
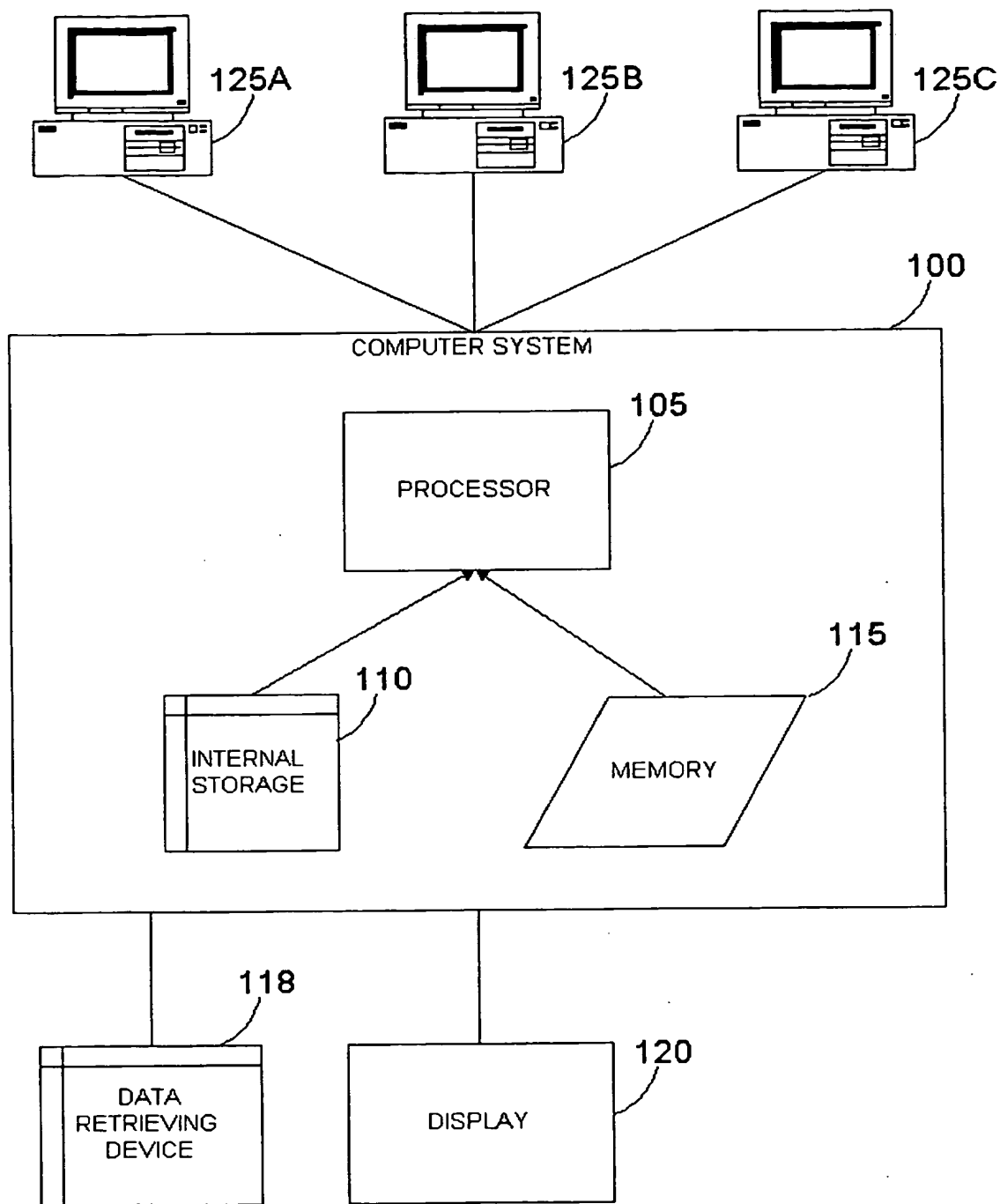


Figure 1

2/5

**Figure 2**

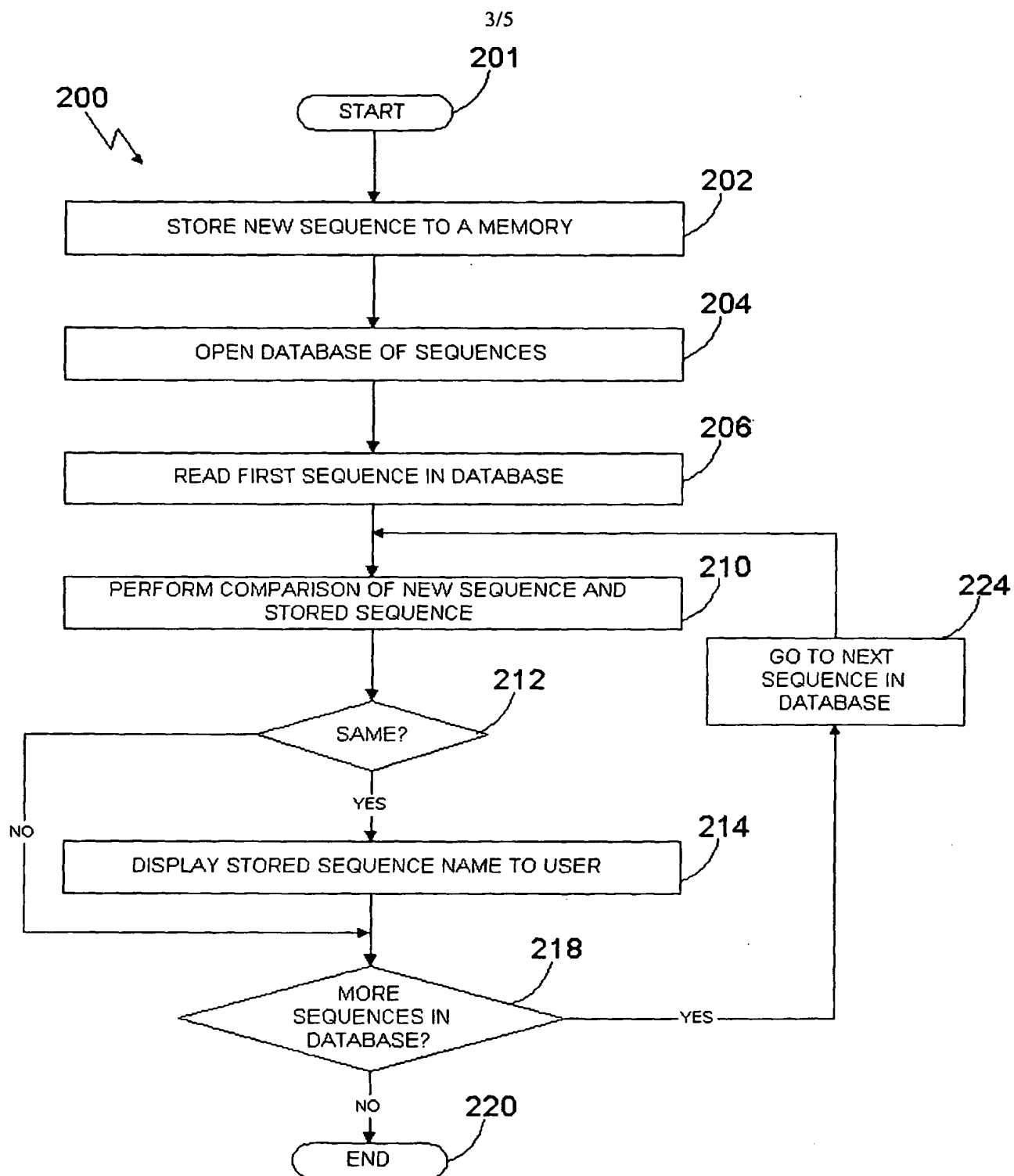


Figure 3

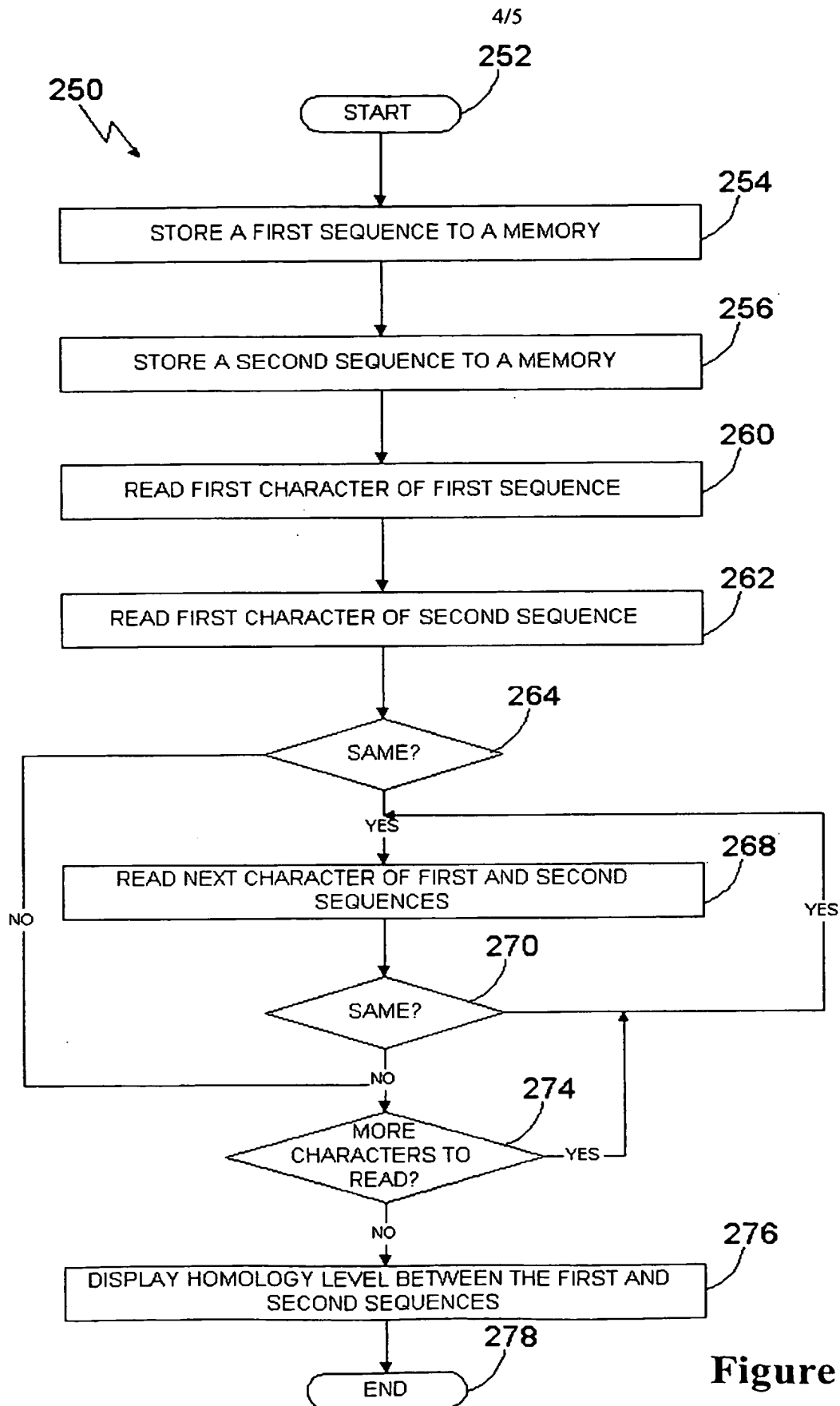
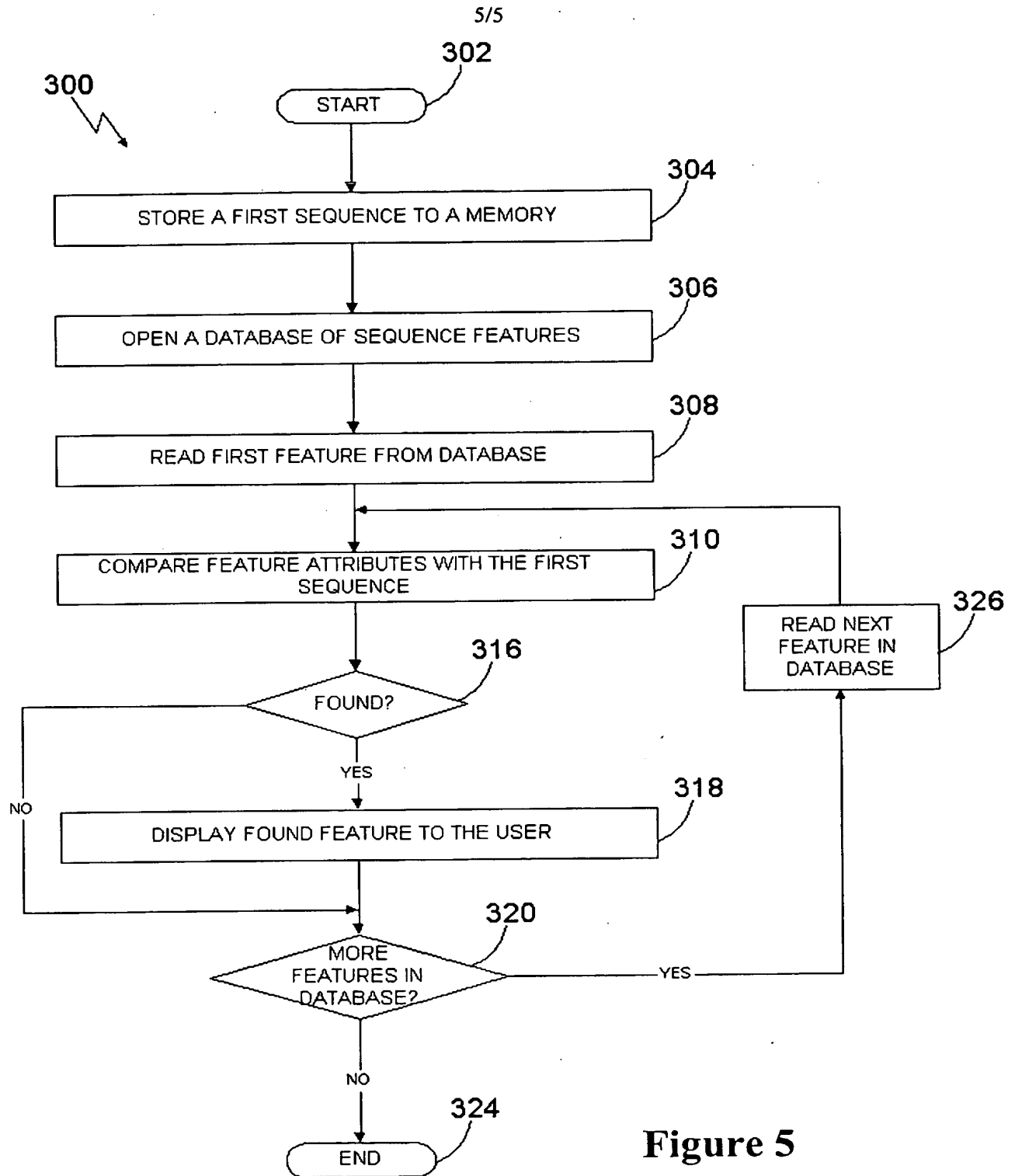


Figure 4

**Figure 5**

<110> GENSET

<120> Full-length human cDNAs encoding potentially secreted proteins

5 <130> 78.WO1

<150> US 60/169,629

<151> 1999-12-08

10 <150> US 60/187,470

<151> 2000-03-06

<160> 482

15 <170> Patent.pm

<210> 1

<211> 2201

<212> DNA

20 <213> Homo sapiens

<220>

<221> CDS

<222> 169..1692

25

<220>

<221> sig_peptide

<222> 169..249

<223> Von Heijne matrix

30 score 7.15265901862021

seq VLLLLLLERGMFS/SP

<400> 1

35 agatgtgaat agctccacta taccagcctc gtcttccttc cgggggacaa cgtgggtcag 60

ggcacagaga gatatttaat gtcaccctct tggggctttc atgggactcc ctctgccaca 120

ttttttggag gttgggaaag ttgctagagg cttcagaact ccagccta atg gat ccc 177

Met Asp Pro

-25

40 aaa ctc ggg aga atg gct gcg tcc ctg ctg gct gtg ctg ctg ctg ctg 225

Lys Leu Gly Arg Met Ala Ala Ser Leu Leu Ala Val Leu Leu Leu Leu

-20 -15 -10

ctg ctg gag cgc ggc atg ttc tcc tca ccc tcc ccg ccc ccg gcg ctg 273

Leu Leu Glu Arg Gly Met Phe Ser Ser Pro Ser Pro Pro Pro Ala Leu

-5 1 5

45 tta gag aaa gtc ttc cag tac att gac ctc cat cag gat gaa ttt gtg 321

Leu Glu Lys Val Phe Gln Tyr Ile Asp Leu His Gln Asp Glu Phe Val

10 15 20

cag acg ctg aag gag tgg gtg gcc atc gag agc gac tct gtc cag cct 369

Gln Thr Leu Lys Glu Trp Val Ala Ile Glu Ser Asp Ser Val Gln Pro

50 25 30 35 40

gtg cct cgc ttc aga caa gag ctc ttc aga atg atg gcc gtg gct gcg 417

Val Pro Arg Phe Arg Gln Glu Leu Phe Arg Met Met Ala Val Ala Ala

45 50 55

gac acg ctg cag cgc ctg ggg gcc cgt gtg gcc tcg gtg gac atg ggt 465

55 Asp Thr Leu Gln Arg Leu Gly Ala Arg Val Ala Ser Val Asp Met Gly

60 65 70

cct cag cag ctg ccc gat ggt cag agt ctt cca ata cct ccc gtc atc 513

Pro Gln Gln Leu Pro Asp Gly Gln Ser Leu Pro Ile Pro Pro Val Ile

75 80 85

60 ctg gcc gaa ctg ggg agc gat ccc acg aaa ggc acc gtg tgc ttc tac 561

Leu Ala Glu Leu Gly Ser Asp Pro Thr Lys Gly Thr Val Cys Phe Tyr

90 95 100

ggc cac ttg gac gtg cag cct gct gac cgg ggc gat ggg tgg ctc acg 609

	Gly	His	Leu	Asp	Val	Gln	Pro	Ala	Asp	Arg	Gly	Asp	Gly	Trp	Leu	Thr	
	105					110					115					120	
	gac	ccc	tat	gtg	ctg	acg	gag	gta	gac	ggg	aaa	ctt	tat	gga	cga	gga	657
5	Asp	Pro	Tyr	Val	Leu	Thr	Glu	Val	Asp	Gly	Lys	Leu	Tyr	Gly	Arg	Gly	
					125					130					135		
	gcg	acc	gac	aac	aaa	ggc	cct	gtc	ttg	gct	tgg	atc	aat	gct	gtg	agc	705
	Ala	Thr	Asp	Asn	Lys	Gly	Pro	Val	Leu	Ala	Trp	Ile	Asn	Ala	Val	Ser	
					140					145					150		
10	gcc	ttc	aga	gcc	ctg	gag	caa	gat	ctt	cct	gtg	aat	atc	aaa	ttc	atc	753
	Ala	Phe	Arg	Ala	Leu	Glu	Gln	Asp	Leu	Pro	Val	Asn	Ile	Lys	Phe	Ile	
					155					160					165		
	att	gag	ggg	atg	gaa	gag	gct	ggc	tct	gtt	gcc	ctg	gag	gaa	ctt	gtg	801
	Ile	Glu	Gly	Met	Glu	Glu	Ala	Gly	Ser	Val	Ala	Leu	Glu	Glu	Leu	Val	
		170					175					180					
15	gaa	aaa	gaa	aag	gac	cga	ttc	ttc	tct	ggt	gtg	gac	tac	att	gta	att	849
	Glu	Lys	Glu	Lys	Asp	Arg	Phe	Phe	Ser	Gly	Val	Asp	Tyr	Ile	Val	Ile	
		185				190					195				200		
	tca	gat	aac	ctg	tgg	atc	agc	caa	agg	aag	cca	gca	atc	act	tat	gga	897
20	Ser	Asp	Asn	Leu	Trp	Ile	Ser	Gln	Arg	Lys	Pro	Ala	Ile	Thr	Tyr	Gly	
					205					210					215		
	acc	cgg	ggg	aac	agc	tac	ttc	atg	gtg	gag	gtg	aaa	tgc	aga	gac	cag	945
	Thr	Arg	Gly	Asn	Ser	Tyr	Phe	Met	Val	Glu	Val	Lys	Cys	Arg	Asp	Gln	
				220						225				230			
25	gat	ttt	cac	tca	gga	acc	ttt	ggt	ggc	atc	ctt	cat	gaa	cca	atg	gct	993
	Asp	Phe	His	Ser	Gly	Thr	Phe	Gly	Gly	Ile	Leu	His	Glu	Pro	Met	Ala	
			235					240					245				
	gat	ctg	gtt	gct	ctt	ctc	ggt	agc	ctg	gta	gac	tcg	tct	ggt	cat	atc	1041
	Asp	Leu	Val	Ala	Leu	Leu	Gly	Ser	Leu	Val	Asp	Ser	Ser	Gly	His	Ile	
		250					255					260					
30	ctg	gtc	cct	gga	atc	tat	gat	gaa	gtg	gtt	cct	ctt	aca	gaa	gag	gaa	1089
	Leu	Val	Pro	Gly	Ile	Tyr	Asp	Glu	Val	Val	Pro	Leu	Thr	Glu	Glu	Glu	
		265				270					275				280		
	ata	aat	aca	tac	aaa	gcc	atc	cat	cta	gac	cta	gaa	gaa	tac	cgg	aat	1137
35	Ile	Asn	Thr	Tyr	Lys	Ala	Ile	His	Leu	Asp	Leu	Glu	Glu	Tyr	Arg	Asn	
					285					290					295		
	agc	agc	cgg	gtt	gag	aaa	ttt	ctg	ttc	gat	act	aag	gag	gag	att	cta	1185
	Ser	Ser	Arg	Val	Glu	Lys	Phe	Leu	Phe	Asp	Thr	Lys	Glu	Glu	Ile	Leu	
				300					305				310				
40	atg	cac	ctc	tgg	agg	tac	cca	tct	ctt	tct	att	cat	ggg	atc	gag	ggc	1233
	Met	His	Leu	Trp	Arg	Tyr	Pro	Ser	Leu	Ser	Ile	His	Gly	Ile	Glu	Gly	
			315					320					325				
	gcg	ttt	gat	gag	cct	gga	act	aaa	aca	gtc	ata	cct	ggc	cga	gtt	ata	1281
	Ala	Phe	Asp	Glu	Pro	Gly	Thr	Lys	Thr	Val	Ile	Pro	Gly	Arg	Val	Ile	
		330					335					340					
45	gga	aaa	ttt	tca	atc	cgt	cta	gtc	cct	cac	atg	aat	gtg	tct	gcg	gtg	1329
	Gly	Lys	Phe	Ser	Ile	Arg	Leu	Val	Pro	His	Met	Asn	Val	Ser	Ala	Val	
		345				350					355				360		
	gaa	aaa	cag	gtg	aca	cga	cat	ctt	gaa	gat	gtg	ttc	tcc	aaa	aga	aat	1377
50	Glu	Lys	Gln	Val	Thr	Arg	His	Leu	Glu	Asp	Val	Phe	Ser	Lys	Arg	Asn	
				365					370						375		
	agt	tcc	aac	aag	atg	gtt	gtt	tcc	atg	act	cta	gga	cta	cac	ccg	tgg	1425
	Ser	Ser	Asn	Lys	Met	Val	Val	Ser	Met	Thr	Leu	Gly	Leu	His	Pro	Trp	
				380					385					390			
55	att	gca	aat	att	gat	gac	acc	cag	tat	ctc	gca	gca	aaa	aga	gcg	atc	1473
	Ile	Ala	Asn	Ile	Asp	Asp	Thr	Gln	Tyr	Leu	Ala	Ala	Lys	Arg	Ala	Ile	
			395					400					405				
	aga	aca	gtg	ttt	gga	aca	gaa	cca	gat	atg	atc	cgg	gat	gga	tcc	acc	1521
	Arg	Thr	Val	Phe	Gly	Thr	Glu	Pro	Asp	Met	Ile	Arg	Asp	Gly	Ser	Thr	
		410					415					420					
60	att	cca	att	gcc	aaa	atg	ttc	cag	gag	atc	gtc	cac	aag	agc	gtg	gtg	1569
	Ile	Pro	Ile	Ala	Lys	Met	Phe	Gln	Glu	Ile	Val	His	Lys	Ser	Val	Val	
		425				430					435				440		
	cta	att	ccg	ctg	gga	gct	gtt	gat	gat	gga	gaa	cat	tcg	cag	aat	gag	1617

Leu Ile Pro Leu Gly Ala Val Asp Asp Gly Glu His Ser Gln Asn Glu
 445 450 455
 aaa atc aac agg tgg aac tac ata gag gga acc aaa tta ttt gct gcc 1665
 Lys Ile Asn Arg Trp Asn Tyr Ile Glu Gly Thr Lys Leu Phe Ala Ala
 460 465 470
 5 ttt ttc tta gag atg gcc cag ctc cat taatcacaag aaccttctag 1712
 Phe Phe Leu Glu Met Ala Gln Leu His
 475 480
 tctgatctga tccactgaca gattcacctc cccacatcc ctagacaggg atggaatgta 1772
 10 aatatccaga gaatttgggt ctagtatagt acattttccc ttccatttaa aatgtcttgg 1832
 gatattctga tcagtaataa aatatttcaa aggcacagat gttggaaatg gtttaaggtc 1892
 cccactgca caccctcctc aagtcatagc tgcttgagc aacttgattt cccaagtcc 1952
 tgtgcaatag ccccaggatt ggattccttc caacctttta gcatactctcc aaccttgcaa 2012
 tttgattggc ataatcactc cagtttgctt tctaggtcct caagtgtctg tgacacataa 2072
 15 tcattccatc caatgatcgc ctttgcttta ccactcttct cttttatctt attaataaaa 2132
 atgttggtct ccaccactga aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaagaaaaaa 2192
 aaaaaaaaaa 2201

 <210> 2
 20 <211> 1631
 <212> DNA
 <213> Homo sapiens

 <220>
 25 <221> CDS
 <222> 148..1140

 <220>
 <221> sig_peptide
 30 <222> 148..240
 <223> Von Heijne matrix
 score 10.0910253445132
 seq LVLLLVTRSPVNA/CL

 35 <400> 2
 gtctgtgcc gccattgtgc ggcgtggtc cctcagagg gttcctgctg ctgccggtgc 60
 cttggaccct cccctcgtct tctcgttcta ctgcccagg agcccggcgg gtccgggact 120
 cccgtccgtg ccggtgcggg cgccggc atg tgg ctg tgg gag gac cag ggc ggc 174
 Met Trp Leu Trp Glu Asp Gln Gly Gly
 -30 -25
 40 ctc ctg ggc cct ttc tcc ttc ctg ctg cta gtg ctg ctg ctg gtg acg 222
 Leu Leu Gly Pro Phe Ser Phe Leu Leu Leu Val Leu Leu Leu Val Thr
 -20 -15 -10
 cgg agc ccg gtc aat gcc tgc ctc ctc acc ggc agc ctc ttc gtt cta 270
 45 Arg Ser Pro Val Asn Ala Cys Leu Leu Thr Gly Ser Leu Phe Val Leu
 -5 1 5 10
 ctg cgc gtc ttc agc ttt gag ccg gtg ccc tct tgc agg gcc ctg cag 318
 Leu Arg Val Phe Ser Phe Glu Pro Val Pro Ser Cys Arg Ala Leu Gln
 15 20 25
 50 gtg ctc aag ccc ccg gac cgc att tct gcc atc gcc cac cgt ggc ggc 366
 Val Leu Lys Pro Arg Asp Arg Ile Ser Ala Ile Ala His Arg Gly Gly
 30 35 40
 agc cac gac gcg ccc gag aac acg ctg gcg gcc att cgg cag gca gct 414
 Ser His Asp Ala Pro Glu Asn Thr Leu Ala Ala Ile Arg Gln Ala Ala
 45 50 55
 aag aat gga gca aca ggc gtg gag ttg gac att gag ttt act tct gac 462
 Lys Asn Gly Ala Thr Gly Val Glu Leu Asp Ile Glu Phe Thr Ser Asp
 60 65 70
 ggg att cct gtc tta atg cac gat aac aca gta gat agg acg act gat 510
 60 Gly Ile Pro Val Leu Met His Asp Asn Thr Val Asp Arg Thr Thr Asp
 75 80 85 90
 ggg act ggg cga ttg tgt gat ttg aca ttt gaa caa att agg aag ctg 558
 Gly Thr Gly Arg Leu Cys Asp Leu Thr Phe Glu Gln Ile Arg Lys Leu

				95				100				105					
	aat	cct	gca	gca	aac	cac	aga	ctc	agg	aat	gat	ttc	cct	gat	gaa	aag	606
	Asn	Pro	Ala	Ala	Asn	His	Arg	Leu	Arg	Asn	Asp	Phe	Pro	Asp	Glu	Lys	
				110				115					120				
5	atc	cct	acc	cta	atg	gaa	gct	gtt	gca	gag	tgc	cta	aac	cat	aac	ctc	654
	Ile	Pro	Thr	Leu	Met	Glu	Ala	Val	Ala	Glu	Cys	Leu	Asn	His	Asn	Leu	
				125				130					135				
	aca	atc	ttc	ttt	gat	gtc	aaa	ggc	cat	gca	cac	aag	gct	act	gag	gct	702
	Thr	Ile	Phe	Phe	Asp	Val	Lys	Gly	His	Ala	His	Lys	Ala	Thr	Glu	Ala	
10				140				145					150				
	cta	aag	aaa	atg	tat	atg	gaa	ttt	cct	caa	ctg	tat	aat	aat	agt	gtg	750
	Leu	Lys	Lys	Met	Tyr	Met	Glu	Phe	Pro	Gln	Leu	Tyr	Asn	Asn	Ser	Val	
				155			160					165				170	
	gtc	tgt	tct	ttc	ttg	cca	gaa	gtt	atc	tac	aag	atg	aga	caa	aca	gat	798
15	Val	Cys	Ser	Phe	Leu	Pro	Glu	Val	Ile	Tyr	Lys	Met	Arg	Gln	Thr	Asp	
					175					180					185		
	cgg	gat	gta	ata	aca	gca	tta	act	cac	aga	cct	tgg	agc	cta	agc	cat	846
	Arg	Asp	Val	Ile	Thr	Ala	Leu	Thr	His	Arg	Pro	Trp	Ser	Leu	Ser	His	
				190					195					200			
20	aca	gga	gat	ggg	aaa	cca	cgc	tat	gat	act	ttc	tgg	aaa	cat	ttt	ata	894
	Thr	Gly	Asp	Gly	Lys	Pro	Arg	Tyr	Asp	Thr	Phe	Trp	Lys	His	Phe	Ile	
				205				210					215				
	ttt	gtt	atg	atg	gac	att	ttg	ctc	gat	tgg	agc	atg	cat	aat	atc	ttg	942
	Phe	Val	Met	Met	Asp	Ile	Leu	Leu	Asp	Trp	Ser	Met	His	Asn	Ile	Leu	
25				220				225					230				
	tgg	tac	ctg	tgt	gga	att	tca	gct	ttc	ctc	atg	caa	aag	gat	ttt	gta	990
	Trp	Tyr	Leu	Cys	Gly	Ile	Ser	Ala	Phe	Leu	Met	Gln	Lys	Asp	Phe	Val	
				235			240				245				250		
	tcc	ccg	gcc	tac	ttg	aag	aag	tgg	tca	gct	aaa	gga	atc	cag	gtt	gtt	1038
30	Ser	Pro	Ala	Tyr	Leu	Lys	Lys	Trp	Ser	Ala	Lys	Gly	Ile	Gln	Val	Val	
					255					260					265		
	ggg	tgg	act	gtt	aat	acc	ttt	gat	gaa	aag	agt	tac	tac	gaa	tcc	cat	1086
	Gly	Trp	Thr	Val	Asn	Thr	Phe	Asp	Glu	Lys	Ser	Tyr	Tyr	Glu	Ser	His	
				270				275						280			
35	ctt	ggg	tcc	agc	tat	atc	act	gac	agc	atg	gta	gaa	gac	tgc	gaa	cct	1134
	Leu	Gly	Ser	Ser	Tyr	Ile	Thr	Asp	Ser	Met	Val	Glu	Asp	Cys	Glu	Pro	
				285				290						295			
	cac	ttc	tagactttca	cggtgggacg	aaacgggttc	agaaactgcc	aggggcctca										1190
	His	Phe															
40				300													
	tacagggata	tcaaaataacc	ctttgtgcta	gcccaggccc	tggggaatca	ggtgactcac											1250
	acaaatgcaa	tagttgggtca	ctgcattttt	acctgaacca	aagctaaacc	cggtgttgcc											1310
	accatgcacc	atggcatgcc	agagttcaac	actgttgctc	ttgaaaatct	ggggtctgaa											1370
	aaaacgcaca	agagcccctg	cctgccccta	gctgaggcac	acagggagac	ccagtgagga											1430
45	taagcacaga	ttgaattgta	caatttgcag	atgcagatgt	aaatgcatgg	gacatgcatg											1490
	ataactcaga	gttgacattt	taaaacttgc	cacacttatt	tcaaatattt	gtactcagct											1550
	atgttaacat	gtactgtaga	catcaaactt	gtggccatac	taataaaaatt	attaaaagga											1610
	gcacaaaaaa	aaaaaaaaaa	a														1631
50	<210> 3																
	<211> 1245																
	<212> DNA																
	<213> Homo sapiens																
55	<220>																
	<221> CDS																
	<222> 85..906																
	<220>																
60	<221> sig_peptide																
	<222> 85..135																
	<223> Von Heijne matrix																
	score 3.86022363031904																

seq GFVAALVAGGVAG/VS

<400> 3

	aaaacatggc	ggcgccagc	gcgcaggac	gtgatccgct	tctgctccgg	cttggattgt	60
5	agccttgacg	aggctctgagc	gacc atg gac cgg ccg ggg ttc gtg gca gcg				111
			Met Asp Arg Pro Gly Phe Val Ala Ala				
			-15			-10	
	ctg gtg gct ggt ggg gta gca ggt gtt tct gtt gac ttg ata tta ttt						159
10	Leu Val Ala Gly Gly Val Ala Gly Val Ser Val Asp Leu Ile Leu Phe						
	-5		1			5	
	cct ctg gat acc att aaa acc agg ctg cag agt ccc caa gga ttt agt						207
	Pro Leu Asp Thr Ile Lys Thr Arg Leu Gln Ser Pro Gln Gly Phe Ser						
	10		15			20	
15	aag gct ggt ggt ttt cat gga ata tat gct ggc gtt cct tct gct gct						255
	Lys Ala Gly Gly Phe His Gly Ile Tyr Ala Gly Val Pro Ser Ala Ala						
	25		30			35	40
	att gga tcc ttt cct aat gct gct gca ttt ttt atc acc tat gaa tat						303
	Ile Gly Ser Phe Pro Asn Ala Ala Ala Phe Phe Ile Thr Tyr Glu Tyr						
	45		50			55	
20	gtg aag tgg ttt ttg cat gct gat tca tct tca tat ttg aca cct atg						351
	Val Lys Trp Phe Leu His Ala Asp Ser Ser Ser Tyr Leu Thr Pro Met						
	60		65			70	
	aaa cat atg ttg gct gcc tct gct gga gaa gtg gtt gcc tgc ctg att						399
25	Lys His Met Leu Ala Ala Ser Ala Gly Glu Val Val Ala Cys Leu Ile						
	75		80			85	
	cga gtt cca tct gaa gtg gtt aag cag agg gca cag gta tct gct tct						447
	Arg Val Pro Ser Glu Val Val Lys Gln Arg Ala Gln Val Ser Ala Ser						
	90		95			100	
	aca aga aca ttt cag att ttc tct aac atc tta tat gaa gag ggt atc						495
30	Thr Arg Thr Phe Gln Ile Phe Ser Asn Ile Leu Tyr Glu Glu Gly Ile						
	105		110			115	120
	caa ggg ttg tat cga ggc tat aaa agc aca gtt tta aga gag att cct						543
	Gln Gly Leu Tyr Arg Gly Tyr Lys Ser Thr Val Leu Arg Glu Ile Pro						
	125		130			135	
35	ttt tct ttg gtc cag ttt ccc tta tgg gag tcc tta aaa gcc ctc tgg						591
	Phe Ser Leu Val Gln Phe Pro Leu Trp Glu Ser Leu Lys Ala Leu Trp						
	140		145			150	
	tcc tgg agg cag gat cat gtg gtg gat tct tgg cag tca gca gtc tgt						639
40	Ser Trp Arg Gln Asp His Val Val Asp Ser Trp Gln Ser Ala Val Cys						
	155		160			165	
	gga gct ttt gca ggt gga ttt gcc gct gca gtc acc acc cct cta gac						687
	Gly Ala Phe Ala Gly Gly Phe Ala Ala Val Thr Thr Pro Leu Asp						
	170		175			180	
	gtg gca aag aca aga att atg ctg gca aag gct ggc tcc agc act gct						735
45	Val Ala Lys Thr Arg Ile Met Leu Ala Lys Ala Gly Ser Ser Thr Ala						
	185		190			195	200
	gat ggg aat gtg ctc tct gtc ctg cat ggg gtc tgg cgg tca cag ggg						783
	Asp Gly Asn Val Leu Ser Val Leu His Gly Val Trp Arg Ser Gln Gly						
	205		210			215	
50	ctg gca gga tta ttt gca ggt gtc ttc cct cga atg gca gcc atc agt						831
	Leu Ala Gly Leu Phe Ala Gly Val Phe Pro Arg Met Ala Ala Ile Ser						
	220		225			230	
	ctg gga ggt ttc atc ttt ctg ggg gct tat gac cga acg cac agc ttg						879
55	Leu Gly Gly Phe Ile Phe Leu Gly Ala Tyr Asp Arg Thr His Ser Leu						
	235		240			245	
	ctg ttg gaa gtt ggc aga aag agt cct tgaagcagag acaagcctca						926
	Leu Leu Glu Val Gly Arg Lys Ser Pro						
	250		255				
	cctccacttc tgtcaagaga ggggcctgca gtgcaaacc						986
60	tgaactatag gccccagtg						1046
	ccatccgtgg tataggctgg ctgggatgaa gtcattggcc						1106
	gaagaaaacg ggggtctgtgg cgggtactctg						1166
	aqtttqgtaa tqctqagaaa aaaaaaaaaa aaaaaaaaaa						1226

aqaaaaaaaaa aaaaaaaaaa

1245

<210> 4

<211> 1623

5 <212> DNA

<213> Homo sapiens

<220>

<221> CDS

10 <222> 31..1248

<220>

<221> sig peptide

<222> 31..135

```
15 <223> Von Heijne matrix
      score 6.3770152988307
      seq TLLLFAAPFGLLG/EK
```

<400> 4

20	aac	cct	tct	tcc	gtc	ggg	ctg	gaa	ttg	cgg	cgc	atg	cgc	ggc	tct	gtg	gag	tgc	acc	54
												Met	Arg	Gly	Ser	Val	Glu	Cys	Thr	
												-35								
	tgg	ggg	tgg	ggg	cac	tgt	gcc	ccc	agc	ccc	ctg	ctc	ctt	tgg	act	cta				102
25	Trp	Gly	Trp	Gly	His	Cys	Ala	Pro	Ser	Pro	Leu	Leu	Leu	Trp	Thr	Leu				
	ctt	ctg	ttt	gca	gcc	cca	ttt	ggc	ctg	ctg	ggg	gag	aag	acc	cgc	cag				150
	Leu	Leu	Phe	Ala	Ala	Pro	Phe	Gly	Leu	Leu	Gly	Glu	Lys	Thr	Arg	Gln				
	gtg	tct	ctg	gag	gtc	atc	cct	aac	tgg	ctg	ggc	ccc	ctg	cag	aac	ctg				198
30	Val	Ser	Leu	Glu	Val	Ile	Pro	Asn	Trp	Leu	Gly	Pro	Leu	Gln	Asn	Leu				
	ctt	cat	ata	cgg	gca	gtg	ggc	acc	aat	tcc	aca	ctg	cac	tat	gtg	tgg				246
	Leu	His	Ile	Arg	Ala	Val	Gly	Thr	Asn	Ser	Thr	Leu	His	Tyr	Val	Trp				
	agc	agc	ctg	ggg	cct	ctg	gca	gtg	gta	atg	gtg	gcc	acc	aac	acc	ccc				294
35	Ser	Ser	Leu	Gly	Pro	Leu	Ala	Val	Val	Met	Val	Ala	Thr	Asn	Thr	Pro				
	cac	agc	acc	ctg	agc	gtc	aac	tgg	agc	ctc	ctg	cta	tcc	cct	gag	ccc				342
40	His	Ser	Thr	Leu	Ser	Val	Asn	Trp	Ser	Leu	Leu	Leu	Ser	Pro	Glu	Pro				
	gat	ggg	ggc	ctg	atg	gtg	ctc	cct	aag	gac	agc	att	cag	ttt	tct	tct				390
	Asp	Gly	Gly	Leu	Met	Val	Leu	Pro	Lys	Asp	Ser	Ile	Gln	Phe	Ser	Ser				
	gcc	ctt	gtt	ttt	acc	agg	ctg	ctt	gag	ttt	gac	agc	acc	aac	gtg	tcc				438
45	Ala	Leu	Val	Phe	Thr	Arg	Leu	Leu	Glu	Phe	Asp	Ser	Thr	Asn	Val	Ser				
	gat	acg	gca	gca	aag	cct	ttg	gga	aga	cca	tat	cct	cca	tac	tcc	ttg				486
	Asp	Thr	Ala	Ala	Lys	Pro	Leu	Gly	Arg	Pro	Tyr	Pro	Pro	Tyr	Ser	Leu				
	gcc	gat	ttc	tct	tgg	aac	aac	atc	act	gat	tca	ttg	gat	cct	gcc	acc				534
50	Ala	Asp	Phe	Ser	Trp	Asn	Asn	Ile	Thr	Asp	Ser	Leu	Asp	Pro	Ala	Thr				
	ctg	agt	gcc	aca	ttt	caa	ggc	cac	ccc	atg	aac	gac	cct	acc	agg	act				582
55	Leu	Ser	Ala	Thr	Phe	Gln	Gly	His	Pro	Met	Asn									

<400> 5
60 gtgtctgcca ctggctgcc ggaggccgaa ggtccctgac tatggctccc cagagcctgc 60
cttcattctag g atg gct cct ctg gcc atg ctg ctt ggg ctg ctg atg gcc 110
Met Ala Pro Leu Gly Met Leu Leu Gly Leu Leu Met Ala
-15 -10 -5

	gcc tgc aca cct tct gcc tca gtc atc aga acc tgaaggagtt tgccttgacc	163
	Ala Cys Thr Pro Ser Ala Ser Val Ile Arg Thr	
	1 5	
	aacccagaga agagcagcac caaagaaacg gagagaaaag aaaccaaagc cgaggaggag	223
5	ctggatgccg aagtcctgga ggtgttccac ccgacgcac agtggcaggc ccttcagcca	283
	gggcaggctg tccctgcagg atcccacgta cggctgaatc ttcagactgg ggaaagagag	343
	gcaaaactcc aatatgagga caagtccga aataatttga aaggcaaaag gctggatatc	403
	aacaccaaca cctacacatc tcaggatctc aagagtgcac tggcaaaatt caaggagggg	463
	gcagagatgg agagttcaaa ggaagacaag gcaaggcagg ctgaggtaaa gcggctcttc	523
10	cgccccattg aggaactgaa gaaagacttt gatgagctga atgttgtcat tgagactgac	583
	atgcagatca tggtagcggc gatcaacaag ttcaatagtt ccagctccag tttggaagag	643
	aagattgctg cgctctttga tcttgaatat tatgtccatc agatggacaa tgcgcaggac	703
	ctgctttcct ttggtggtct tcaagtgggtg atcaatgggc tgaacagcac agagccctc	763
	gtgaaggagt atgctgcgtt tgtgctgggc gctgcctttt ccagcaaccc caaggtccag	823
15	gtggaggcca tcgaaggggg agccctgcag aagctgctgg tcatcctggc cacggagcag	883
	ccgctcactg caaagggagg tgctcaccgt gcgcgtgggc acactgctct acgacctggt	943
	cacggagaag atgttcgccc aggaggaggc tgagctgacc caggagatgt ccccagagaa	1003
	gctgcagcag tatcgccagg tacacctcct gccakgcctg tgggaacagg gctgggtgca	1063
	gatcacggcc cacctcctgg cgctgcccga gcatgatgcc ygtgagaagg tgctgcwgac	1123
20	actgggcgtc ctccctgacca cctgcccggg ccgctaccgt caggaccccc agctcggcag	1183
	gacactggcc agcctgcagg ctgagtacca ggtgctggcc agcctggagc tgcaggatgg	1243
	tgaggacgag ggctacttcc aggagctgct gggctctgtc aacagcttgc tgaaggagct	1303
	gagatgaggc cccacaccag gactggactg ggatgccgct agtgaggctg aggggtgcca	1363
	gcgtgggtgg gcttctcagg caggaggaca tcttggcagt gctggcttgg ccattaaatg	1423
25	gaaacctgaa ggccaaaaaa aaaaaaaaaa a	1454
	<210> 6	
	<211> 1639	
	<212> DNA	
30	<213> Homo sapiens	
	<220>	
	<221> CDS	
	<222> 111..1154	
35	<220>	
	<221> sig_peptide	
	<222> 111..197	
	<223> Von Heijne matrix	
40	score 4.68065944212013	
	seq LLGPLMAACFTFC/LS	
	<400> 6	
	agacggctgc cgccgcgttt gcgcaggggg agctggctgc cgccgcggcc gcctggaatt	60
45	gtgggagttg tgtctgccac tcggctgccc gagggccgaag gtccctgact atg gct	116
	Met Ala	
	ccc cag agc ctg cct tca tct agg atg gct cct ctg ggc atg ctg ctt	164
	Pro Gln Ser Leu Pro Ser Ser Arg Met Ala Pro Leu Gly Met Leu Leu	
	-25 -20 -15	
50	ggg ccg ctg atg gcc gcc tgc ttc acc ttc tgc ctc agt cat cag aac	212
	Gly Pro Leu Met Ala Ala Cys Phe Thr Phe Cys Leu Ser His Gln Asn	
	-10 -5 1 5	
	ctg aag gag ttt gcc ctg acc aac cca gag aag agc agc acc aaa gaa	260
	Leu Lys Glu Phe Ala Leu Thr Asn Pro Glu Lys Ser Ser Thr Lys Glu	
55	10 15 20	
	aca gag aga aaa gaa acc aaa gcc gag gag gag ctg gat gcc gaa gtc	308
	Thr Glu Arg Lys Glu Thr Lys Ala Glu Glu Glu Leu Asp Ala Glu Val	
	25 30 35	
	ctg gag gtg ttc cac ccg acg cat gag tgg cag gcc ctt cag cca ggg	356
60	Leu Glu Val Phe His Pro Thr His Glu Trp Gln Ala Leu Gln Pro Gly	
	40 45 50	
	cag gct gtc cct gca gga tcc cac gta cgg ctg aat ctt cag act ggg	404
	Gln Ala Val Pro Ala Gly Ser His Val Arg Leu Asn Leu Gln Thr Gly	

	55		60		65		
	gaa aga gag gca aaa ctc caa tat gag gac aag ttc cga aat aat ttg						452
	Glu Arg Glu Ala Lys Leu Gln Tyr Glu Asp Lys Phe Arg Asn Asn Leu						
	70		75		80		85
5	aaa ggc aaa agg ctg gat atc aac acc aac acc tac aca tct cag gat						500
	Lys Gly Lys Arg Leu Asp Ile Asn Thr Asn Thr Tyr Thr Ser Gln Asp						
		90		95			100
	ctc aag agt gca ctg gca aaa ttc aag gag ggg gca gag atg gag agt						548
10	Leu Lys Ser Ala Leu Ala Lys Phe Lys Glu Gly Ala Glu Met Glu Ser						
		105		110			115
	tca aag gaa gac aag gca agg cag gct gag gta aag cgg ctc ttc cgc						596
	Ser Lys Glu Asp Lys Ala Arg Gln Ala Glu Val Lys Arg Leu Phe Arg						
		120		125			130
	ccc att gag gaa ctg aag aaa gac ttt gat gag ctg aat gtt gtc att						644
15	Pro Ile Glu Glu Leu Lys Lys Asp Phe Asp Glu Leu Asn Val Val Ile						
		135		140			145
	gag act gac atg cag atc atg gta cgg ctg atc aac aag ttc aat agt						692
	Glu Thr Asp Met Gln Ile Met Val Arg Leu Ile Asn Lys Phe Asn Ser						
		150		155			160
20	tcc agc tcc agt ttg gaa gag aag att gct gcg ctc ttt gat ctt gaa						740
	Ser Ser Ser Ser Leu Glu Glu Lys Ile Ala Ala Leu Phe Asp Leu Glu						
		170		175			180
	tat tat gtc cat cag atg gac aat gcg cag gac ctg ctt tcc ttt ggt						788
25	Tyr Tyr Val His Gln Met Asp Asn Ala Gln Asp Leu Leu Ser Phe Gly						
		185		190			195
	ggt ctt caa gtg gtg atc aat ggg ctg aac agc aca gag ccc ctc gtg						836
	Gly Leu Gln Val Val Ile Asn Gly Leu Asn Ser Thr Glu Pro Leu Val						
		200		205			210
	aag gag tat gct gcg ttt gtg ctg ggc gct gcc ttt tcc agc aac ccc						884
30	Lys Glu Tyr Ala Ala Phe Val Leu Gly Ala Ala Phe Ser Ser Asn Pro						
		215		220			225
	aag gtc cag gtg gag gcc atc gaa ggg gga gcc ctg cag aag ctg ctg						932
	Lys Val Gln Val Glu Ala Ile Glu Gly Gly Ala Leu Gln Lys Leu Leu						
		230		235			240
35	gtc atc ctg gcc acg gag cag ccg ctc act gca aag aag aag gtc ctg						980
	Val Ile Leu Ala Thr Glu Gln Pro Leu Thr Ala Lys Lys Lys Val Leu						
		250		255			260
	ttt gca ctg tgc tcc ctg ctg cgc cac ttc ccc tat gcc cag cgg cag						1028
40	Phe Ala Leu Cys Ser Leu Leu Arg His Phe Pro Tyr Ala Gln Arg Gln						
		265		270			275
	ttc ctg aag ctc ggg ggg ctg cag gtc ctg agg acc ctg gtg cag gag						1076
	Phe Leu Lys Leu Gly Gly Leu Gln Val Leu Arg Thr Leu Val Gln Glu						
		280		285			290
	aag ggc acg gag gtg ctc gcc gtg cgc gtg gtc aca ctg ctc tac gac						1124
45	Lys Gly Thr Glu Val Leu Ala Val Arg Val Val Thr Leu Leu Tyr Asp						
		295		300			305
	ctg gtc acg gag aag atg ttc gcc gag gag taggctgagc tgacccagga						1174
	Leu Val Thr Glu Lys Met Phe Ala Glu Glu						
		310		315			
50	gatgtcccca gagaagctgc agcagtatcg ccaggtacac ctccctgccag gcctgtggga						1234
	acaggggctgg tgcgagatca cggccacact cctggcgtcg cccgagcatg atgccgtga						1294
	gaaggtgctg cagacactgg gcgtcctcct gaccacctgc cgggaccgct accgtcagga						1354
	ccccagctc ggcaggacac tggccagcct gcaggctgag taccaggtgc tggccagcct						1414
	ggagctgcag gatggtgagg acgaggggcta cttccaggag ctgctgggct ctgtcaacag						1474
55	cttgctgaag gagctgagat gagggccac accaggactg gactgggatg ccgctagtga						1534
	ggctgagggg tgccagcgtg ggtgggcttc tcaggcagga ggacatcttg gcagtgtctgg						1594
	cttggccatt aaatggaac ctgaaggcaa aaaaaaaaaa aaaaa						1639

<210> 7

60 <211> 1768

<212> DNA

<213> Homo sapiens

```

<220>
<221> CDS
<222> 66..1256

5 <220>
  <221> sig_peptide
  <222> 66..173
  <223> Von Heijne matrix
      score 4.89555877630516
10      seq LLLRLNDAALRA/LQ

<400> 7
agaggaggtg gcggtggtgg ccctcgcttg tggccccgt gctgcttgca ctcgaactcg      60
tcgcc atg gag gag ctc cag gag cct ctg aga gga cag ctc cgg ctc tgc      110
15      Met Glu Glu Leu Gln Glu Pro Leu Arg Gly Gln Leu Arg Leu Cys
      -35 -30 -25
      ttc acg caa gct gcc cgg act agc ctc tta ctg ctc agg ctc aac gac      158
      Phe Thr Gln Ala Ala Arg Thr Ser Leu Leu Leu Leu Arg Leu Asn Asp
      -20 -15 -10
20 gct gcc ctg cgg gcg ctg caa gag tgt cag cgg caa cag gta cgg ccg      206
      Ala Ala Leu Arg Ala Leu Gln Glu Cys Gln Arg Gln Gln Val Arg Pro
      -5 1 5 10
      gtg att gct ttc caa ggc cac cga ggg tat ctg aga ctc cca ggc cct      254
      Val Ile Ala Phe Gln Gly His Arg Gly Tyr Leu Arg Leu Pro Gly Pro
      15 20 25
25 ggt tgg tcc tgc ctc ttc tcc ttc ata gtg tcc cag tgt tgt cag gag      302
      Gly Trp Ser Cys Leu Phe Ser Phe Ile Val Ser Gln Cys Cys Gln Glu
      30 35 40
      ggc gct ggt ggt agc ttg gac ctt gtg tgc caa cgc ttc ctc agg tct      350
      Gly Ala Gly Gly Ser Leu Asp Leu Val Cys Gln Arg Phe Leu Arg Ser
      45 50 55
      ggg cct aac agc ctc cac tgc ctg ggc tca ctc agg gag cgc ctc att      398
      Gly Pro Asn Ser Leu His Cys Leu Gly Ser Leu Arg Glu Arg Leu Ile
      60 65 70 75
35 att tgg gca gcc atg gat tct atc cca gcc cca tca tca gtt cag gga      446
      Ile Trp Ala Ala Met Asp Ser Ile Pro Ala Pro Ser Ser Val Gln Gly
      80 85 90
      cac aac ctg act gaa gat gcc aga cat cct gag agt tgg cag aac aca      494
      His Asn Leu Thr Glu Asp Ala Arg His Pro Glu Ser Trp Gln Asn Thr
      95 100 105
40 gga ggc tat tct gaa gga gat gca gta tca cag cca cag atg gca cta      542
      Gly Gly Tyr Ser Glu Gly Asp Ala Val Ser Gln Pro Gln Met Ala Leu
      110 115 120
      gag gag gtg tca gtg tca gat cca ctg gca agc aac caa gga cag tca      590
      Glu Glu Val Ser Val Ser Asp Pro Leu Ala Ser Asn Gln Gly Gln Ser
      125 130 135
      ctc cca gga tcc tca agg gag cac atg gca cag tgg gaa gtg aga agc      638
      Leu Pro Gly Ser Ser Arg Glu His Met Ala Gln Trp Glu Val Arg Ser
      140 145 150 155
50 cag acc cat gtt cca aac aga gaa cct gtt cag gca ctg cct tcc tct      686
      Gln Thr His Val Pro Asn Arg Glu Pro Val Gln Ala Leu Pro Ser Ser
      160 165 170
      gcc agc cgg aaa cgt ctg gac aag aaa cgt tca gtg cct gta gcc act      734
      Ala Ser Arg Lys Arg Leu Asp Lys Lys Arg Ser Val Pro Val Ala Thr
      175 180 185
55 gta gaa ctg gaa gaa aag agg ttc aga act ctg cct tta gtg cca agc      782
      Val Glu Leu Glu Glu Lys Arg Phe Arg Thr Leu Pro Leu Val Pro Ser
      190 195 200
      ccc cta caa ggc ctg acc aat cag gat tta caa gag gga gaa gat tgg      830
      Pro Leu Gln Gly Leu Thr Asn Gln Asp Leu Gln Glu Gly Glu Asp Trp
      205 210 215
      gag caa gaa gat gag gac atg gac ccc aga tta gaa cac agt tcc tca      878
      Glu Gln Glu Asp Glu Asp Met Asp Pro Arg Leu Glu His Ser Ser Ser

```

	220		225		230		235	
	ggt caa gaa gat tct gaa tcc cca agt cct gaa gat ata cca gac tac	926						
	Val Gln Glu Asp Ser Glu Ser Pro Ser Pro Glu Asp Ile Pro Asp Tyr							
		240		245		250		
5	ctc ctg caa tac agg gcc atc cac agt gca gaa cag caa cat gcc tat	974						
	Leu Leu Gln Tyr Arg Ala Ile His Ser Ala Glu Gln Gln His Ala Tyr							
		255		260		265		
	gag cag gac ttt gag aca gat tat gct gaa tac cgc atc ctg cat gcc	1022						
	Glu Gln Asp Phe Glu Thr Asp Tyr Ala Glu Tyr Arg Ile Leu His Ala							
10		270		275		280		
	cgt gtt ggg act gca agc caa agg ttc ata gag ctg gga gca gag att	1070						
	Arg Val Gly Thr Ala Ser Gln Arg Phe Ile Glu Leu Gly Ala Glu Ile							
		285		290		295		
	aaa aga gtt cgg cga gga act cca gaa tac aag gtc ctg gaa gac aag	1118						
15	Lys Arg Val Arg Arg Gly Thr Pro Glu Tyr Lys Val Leu Glu Asp Lys							
		300		305		310		315
	ata atc cag gaa tat aaa aag ttc agg aag cag tac cca agt tac aga	1166						
	Ile Ile Gln Glu Tyr Lys Lys Phe Arg Lys Gln Tyr Pro Ser Tyr Arg							
		320		325		330		
20	gaa gaa aag cgt cgc tgt gag tac ctt cac cag aaa ttg tcc cac att	1214						
	Glu Glu Lys Arg Arg Cys Glu Tyr Leu His Gln Lys Leu Ser His Ile							
		335		340		345		
	aaa ggt ctc atc ctg gag ttt gag gaa aag aac agg ggc agc	1256						
	Lys Gly Leu Ile Leu Glu Phe Glu Glu Lys Asn Arg Gly Ser							
25		350		355		360		
	tgaagttatc aagggaattt ttgagcctct gcttagtgaa acacaaagga acaaagcagc	1316						
	tataaactaa atagaatgca actatctgct tttcttatgc tgaccactgg agtccatggg	1376						
	ggcaagtaga gagctgctct aggttcttga ggtttggttt tcattattaa tttttagggt	1436						
	atgggcactg tgcaaaagact ccatagctgt gcctaggagt ctaggaaaag tgacagaggc	1496						
30	ttggcttttt tacttttagt tcagccaagt catttttcaag tcctgagaaa tgacatcatc	1556						
	ttcaggataa aataatgagg acattagaca aaccaaacta agtgaatttt agcctggtag	1616						
	cctctctaag gaaacagtaa taataacttc tgataagagt taaaagaact tgtagcatac	1676						
	ctggatataa tgggaaaagg cctgggtgtt acccatgtac tgaaaatgaa cttttaccaa	1736						
	catggctaaa aaattaaaaa aaaaaaaaaa aa	1768						
35								
	<210> 8							
	<211> 1510							
	<212> DNA							
	<213> Homo sapiens							
40								
	<220>							
	<221> CDS							
	<222> 190..1398							
45								
	<220>							
	<221> sig_peptide							
	<222> 190..252							
	<223> Von Heijne matrix							
	score 5.8172934575094							
50	seq ALLWAQEVGVLA/GR							
	<400> 8							
	acggttgccc tggcagcgcg cgaggctggt gagtcggcag ccctgtggca gccggcgggc	60						
	tggtttccat ggttgcacga ttaggaacca ccagctgctg catcccatgg ccaggggtgg	120						
55	cgtccaggtg gcagagcagc taggaacgca aggcctgaac ctggggccag acaccctgct	180						
	ctcccggcc atg gtc aac gac cct cca gta cct gcc tta ctg tgg gcc cag	231						
	Met Val Asn Asp Pro Pro Val Pro Ala Leu Leu Trp Ala Gln							
	-20 -15 -10							
	gag gtg ggc caa gtc ttg gca ggc cgt gcc cgc agg ctg ctg ctg cag	279						
60	Glu Val Gly Gln Val Leu Ala Gly Arg Ala Arg Arg Leu Leu Leu Gln							
	-5 1 5							
	ttt ggg gtg ctc ttc tgc acc atc ctc ctt ttg ctc tgg gtg tct gtc	327						
	Phe Gly Val Leu Phe Cys Thr Ile Leu Leu Leu Leu Trp Val Ser Val							

	10					15					20				25		
	ttc	ctc	tat	ggc	tcc	ttc	tac	tat	tcc	tat	atg	ccg	aca	gtc	agc	cac	375
	Phe	Leu	Tyr	Gly	Ser	Phe	Tyr	Tyr	Ser	Tyr	Met	Pro	Thr	Val	Ser	His	
					30					35					40		
5	ctc	agc	cct	gtg	cat	ttc	tac	tac	agg	acc	gac	tgt	gat	tcc	tcc	acc	423
	Leu	Ser	Pro	Val	His	Phe	Tyr	Tyr	Arg	Thr	Asp	Cys	Asp	Ser	Ser	Thr	
				45					50					55			
	acc	tca	ctc	tgc	tcc	ttc	cct	gtt	gcc	aat	gtc	tcg	ctg	act	aag	ggt	471
	Thr	Ser	Leu	Cys	Ser	Phe	Pro	Val	Ala	Asn	Val	Ser	Leu	Thr	Lys	Gly	
10			60					65					70				
	gga	cgt	gat	cgg	gtg	ctg	atg	tat	gga	cag	ccg	tat	cgt	gtt	acc	tta	519
	Gly	Arg	Asp	Arg	Val	Leu	Met	Tyr	Gly	Gln	Pro	Tyr	Arg	Val	Thr	Leu	
		75					80					85					
	gag	ctt	gag	ctg	cca	gag	tcc	cct	gtg	aat	caa	gat	ttg	ggc	atg	ttc	567
15	Glu	Leu	Glu	Leu	Pro	Glu	Ser	Pro	Val	Asn	Gln	Asp	Leu	Gly	Met	Phe	
						95					100				105		
	ttg	gtc	acc	att	tcc	tgc	tac	acc	aga	ggt	ggc	cga	atc	atc	tcc	act	615
	Leu	Val	Thr	Ile	Ser	Cys	Tyr	Thr	Arg	Gly	Gly	Arg	Ile	Ile	Ser	Thr	
					110					115					120		
20	tct	tcg	cgt	tcg	gtg	atg	ctg	cat	tac	cgc	tca	gac	ctg	ctc	cag	atg	663
	Ser	Ser	Arg	Ser	Val	Met	Leu	His	Tyr	Arg	Ser	Asp	Leu	Leu	Gln	Met	
				125					130					135			
	ctg	gac	aca	ctg	gtc	ttc	tct	agc	ctc	ctg	cta	ttt	ggc	ttt	gca	gag	711
	Leu	Asp	Thr	Leu	Val	Phe	Ser	Ser	Leu	Leu	Leu	Phe	Gly	Phe	Ala	Glu	
25			140					145					150				
	cag	aag	cag	ctg	ctg	gag	gtg	gaa	ctc	tac	gca	gac	tat	aga	gag	aac	759
	Gln	Lys	Gln	Leu	Leu	Glu	Val	Glu	Leu	Tyr	Ala	Asp	Tyr	Arg	Glu	Asn	
		155					160					165					
	tcg	gtg	agt	gag	tac	gtg	ccg	acc	act	gga	gcg	atc	att	gag	atc	cac	807
30	Ser	Val	Ser	Glu	Tyr	Val	Pro	Thr	Thr	Gly	Ala	Ile	Ile	Glu	Ile	His	
						175					180					185	
	agc	aag	cgc	atc	cag	ctg	tat	gga	gcc	tac	ctc	cgc	atc	cac	gcg	cac	855
	Ser	Lys	Arg	Ile	Gln	Leu	Tyr	Gly	Ala	Tyr	Leu	Arg	Ile	His	Ala	His	
				190					195						200		
35	ttc	act	ggg	ctc	aga	tac	ctg	cta	tac	aac	ttc	ccg	atg	acc	tgc	gcc	903
	Phe	Thr	Gly	Leu	Arg	Tyr	Leu	Leu	Tyr	Asn	Phe	Pro	Met	Thr	Cys	Ala	
				205					210					215			
	ttc	ata	ggg	gtt	gcc	agc	aac	ttc	acc	ttc	ctc	agc	gtc	atc	gtg	ctc	951
	Phe	Ile	Gly	Val	Ala	Ser	Asn	Phe	Thr	Phe	Leu	Ser	Val	Ile	Val	Leu	
40			220					225					230				
	ttc	agc	tac	atg	cag	tgg	gtg	tgg	ggg	ggc	atc	tgg	ccc	cga	cac	cgc	999
	Phe	Ser	Tyr	Met	Gln	Trp	Val	Trp	Gly	Gly	Ile	Trp	Pro	Arg	His	Arg	
		235					240					245					
	ttc	tct	ttg	cag	gtt	aac	atc	cga	aaa	aga	gac	aat	tcc	cgg	aag	gaa	1047
45	Phe	Ser	Leu	Gln	Val	Asn	Ile	Arg	Lys	Arg	Asp	Asn	Ser	Arg	Lys	Glu	
						255					260					265	
	gtc	caa	cga	agg	atc	tct	gct	cat	cag	cca	ggg	gca	ggg	cct	gaa	ggc	1095
	Val	Gln	Arg	Arg	Ile	Ser	Ala	His	Gln	Pro	Gly	Ala	Gly	Pro	Glu	Gly	
				270						275					280		
50	cag	gag	gag	tca	act	ccg	caa	tca	gat	gtt	aca	gag	gat	ggg	gag	agc	1143
	Gln	Glu	Glu	Ser	Thr	Pro	Gln	Ser	Asp	Val	Thr	Glu	Asp	Gly	Glu	Ser	
				285					290					295			
	cct	gaa	gat	ccc	tca	ggg	aca	gag	ggg	cag	ctg	tcc	gag	gag	gag	aaa	1191
	Pro	Glu	Asp	Pro	Ser	Gly	Thr	Glu	Gly	Gln	Leu	Ser	Glu	Glu	Glu	Lys	
55			300					305					310				
	cca	gat	cag	cag	ccc	ctg	agc	gga	gaa	gag	gag	cta	gag	cct	gag	gcc	1239
	Pro	Asp	Gln	Gln	Pro	Leu	Ser	Gly	Glu	Glu	Glu	Leu	Glu	Pro	Glu	Ala	
		315					320					325					
	agt	gat	ggg	tca	ggc	tcc	tgg	gaa	gat	gca	gct	ttg	ctg	acg	gag	gcc	1287
60	Ser	Asp	Gly	Ser	Gly	Ser	Trp	Glu	Asp	Ala	Ala	Leu	Leu	Thr	Glu	Ala	
		330				335					340					345	
	aac	ctg	cct	gct	cct	gct	cct	gct	tct	gct	tct	gcc	cct	gtc	cta	gag	1335
	Asn	Leu	Pro	Ala	Pro	Ala	Pro	Ala	Ser	Ala	Ser	Ala	Pro	Val	Leu	Glu	

```

          350          355          360
act ctg ggc agc tct gaa cct gct ggg ggt gct ctc cga cag cgc ccc 1383
Thr Leu Gly Ser Ser Glu Pro Ala Gly Gly Ala Leu Arg Gln Arg Pro
          365          370          375
5 acc tgc tct agt tcc tgaagaaaag gggcagactc ctcacattcc agcactttcc 1438
Thr Cys Ser Ser Ser
          380
cacctgactc ctctccctc gtttttcctt caataaacta ttttgtgtca gctccaaaaa 1498
aaaaaaaaaa aa 1510
10
<210> 9
<211> 882
<212> DNA
<213> Homo sapiens
15
<220>
<221> CDS
<222> 78..410
20
<220>
<221> sig_peptide
<222> 78..155
<223> Von Heijne matrix
      score 10.0731536331164
25      seq LWLALVSCILTQA/SA

<400> 9
atggctggcc agaggaggaa cgctttgtgt tctcatcgga gctgcatggg aagtctgcat 60
acagcaaaagt gacctgc atg cct cac ctt atg gaa agg atg gtg ggc tct 110
30      Met Pro His Leu Met Glu Arg Met Val Gly Ser
          -25          -20
ggc ctc ctg tgg ctg gcc ttg gtc tcc tgc att ctg acc cag gca tct 158
Gly Leu Leu Trp Leu Ala Leu Val Ser Cys Ile Leu Thr Gln Ala Ser
-15          -10          -5          1
35 gca gtg cag cga ggt tat gga aac ccc att gaa gcc agt tgc tat ggg 206
Ala Val Gln Arg Gly Tyr Gly Asn Pro Ile Glu Ala Ser Ser Tyr Gly
          5          10          15
ctg gac ctg gac tgc gga gct cct ggc acc cca gag gct cat gtc tgt 254
Leu Asp Leu Asp Cys Gly Ala Pro Gly Thr Pro Glu Ala His Val Cys
40      20      25      30
ttt gac ccc tgt cag aat tac acc ctc cta gat ttg ggg ccc atc act 302
Phe Asp Pro Cys Gln Asn Tyr Thr Leu Leu Asp Leu Gly Pro ile Thr
35      40      45
cgg aga ggt gca cag tct ccc ggt gtc atg aat gga acc cct agc act 350
Arg Arg Gly Ala Gln Ser Pro Gly Val Met Asn Gly Thr Pro Ser Thr
50      55      60      65
gca ggg ttc ctg gtg gcc tgg cct atg gtc ctc ctg act gtc ctc ctg 398
Ala Gly Phe Leu Val Ala Trp Pro Met Val Leu Leu Thr Val Leu Leu
          70      75      80
50 gct tgg ctg ttc tgagagctcc gctgagcatc tggccttgaa gtttgtgttc 450
Ala Trp Leu Phe
          85
ttccctctgg caatggctcc cttcagcact tctgctttcc actccaattc acacaggctt 510
ggtattaaca gaatcaaggc caggctaggt taggaaaagg gaagagcttt caccttcttt 570
55 aaaactctcg gctgggcgca gtggctcatg cctgtaatcc cagcattttg ggaggctgag 630
gcaggtggat cacctgaggt cagcagttca aaatcagcct ggccaaaatg ctgaaactcc 690
gtctctacta aaaatacaaa aattagccag gcatgggtgac aggcgcctgt aatcccagct 750
actcgggagg ccaaggcagg agaattgctc gaactcaggg ggtggagggt gcagtgagtt 810
gagattgtgc cattgcactc cagcctgggc aacagagcaa gactctgtct caggcaaaaa 870
60 aaaaaaaaaa aa 882

<210> 10
<211> 1849

```

<212> DNA
 <213> Homo sapiens

 <220>
 5 <221> CDS
 <222> 84..299

 <220>
 <221> sig_peptide
 10 <222> 84..134
 <223> Von Heijne matrix
 score 3.86022363031904
 seq GFVAALVAGGVAG/VS

 15 <400> 10
 aaacatggcg gcgcccagcg cgcgaggacg tgatccgctt ctgctccggc ttggattgta 60
 gccttgacga ggtctgagcg acc atg gac cgg ccg ggg ttc gtg gca gcg ctg 113
 Met Asp Arg Pro Gly Phe Val Ala Ala Leu
 -15 -10
 20 gtg gct ggt ggg gta gca ggt gtt tct gtt gac ttg ata tta ttt cct 161
 Val Ala Gly Gly Val Ala Gly Val Ser Val Asp Leu Ile Leu Phe Pro
 -5 1 5
 ctg gat acc att aaa acc agg ctg cag agt ccc caa gga ttt aat aag 209
 Leu Asp Thr Ile Lys Thr Arg Leu Gln Ser Pro Gln Gly Phe Asn Lys
 25 10 15 20 25
 gct ggt ggt ttt cat gga ata tat gct ggc gtt cct tct gct gct att 257
 Ala Gly Gly Phe His Gly Ile Tyr Ala Gly Val Pro Ser Ala Ala Ile
 30 35 40
 gga tcc ttt cct aat ggt tgc ctg cct gat tgc agt tcc atc 299
 Gly Ser Phe Pro Asn Gly Cys Leu Pro Asp Ser Ser Ser Ile
 45 50 55
 tgaagtgggt aagcagaggg cacaggtatc tgcttctaca agaacatttc agattttctc 359
 taacatctta tatgaagagg gtatccaagg gttgtatcga ggctataaaa gcacagtttt 419
 aagagagatt cctttttctt tgggtccagtt tcccttatgg gagtccttaa aagccctctg 479
 35 gtcttgagg caggatcatg tgggtgattc ttggcagtcg gcagtcctgtg gagcttttgc 539
 aggtggattt gccgctgcag tcaccacccc tctagacgtg gcgaagacaa gaattatgct 599
 ggcaaaaggct ggctccagca ctgctgatgg gaatgtgctc tctgtcctgc atggggctctg 659
 gcggtcacag gggctggcag gattatttgc aggtgtcttc cctcgaatgg cagccatcag 719
 tctgggagggt ttcattctttc tgggggctta tgaccgaacg cacagcttgc tgttggaggt 779
 40 tggcagaaaag agtccttgaa gcagagacaa gcctcacctc cacttctgtc aagagagggg 839
 cctgcagtggt aaacctctct ccgctgagca gctgtctgaa ctataggccc cagtgtgtaa 899
 gaccagttgt gctaagatac cggcatggag attgtgccat ccgtgggtata ggctggctgg 959
 tatgaagtca ttggcctgta tgccagagag ctaagagaag aaaacgggggt ctgtggcagt 1019
 actctgaaca atttctctcag aacctcttaa taaataagtt tggtaatgct gaggccaggc 1079
 45 ctttttagagc tttcatttga tctgtatctg atctttcatt tcctgccacc tgatggtgga 1139
 ttcagcagaa ggcaagatgg ttataattct aaaagaatag cttgtttgtt tgtttgtttg 1199
 ggaaaaggag acttggggaa gagttgtgta tgtgggtgtt tctcccccta gttaattcct 1259
 gttgtgtaag ggtaggcttt gttgaaaaag aaagaaagat tgaactacag gtgcatagca 1319
 agcactcttt ctgggtaact aggctgctgg ttttaattac cctcagattt caccataaaa 1379
 50 aacgcacaat tgtattattt tacagagatg tgtccagcgc cccctgtggg gtgtgagaga 1439
 aagcagctgc aactcaagtg actaggtggg cccagctggc ttcgtgcagg agggcacggg 1499
 ggggtgagcca ttctcgccat tctcatgtca gactgaaagg agggcctggg ccagctttga 1559
 aaaggcagga tgaaatggaa aggtcaccac acttagggat tttagacctt gactaacaag 1619
 ctccaggtgt agaaaaattc aaaacaaaat gtcaggaatc tagcagtggt gtctgccctg 1679
 55 gagcaaaaca acagtatgtg attttgcttc gcctattttt tttttctttt ttgggggaag 1739
 ataattaaag gcagaatgac tgcgtttgta aaagaaggac caccaactat actgacattt 1799
 ataaatgaac ctttattaaa gacacttcaa tgcaaaaaaa aaaaaaaaaa 1849

 <210> 11
 60 <211> 565
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 55..468

5 <220>
 <221> sig_peptide
 <222> 55..99
 <223> Von Heijne matrix
 score 8.96936032049195
 10 seq FTLLFLAAVAGA/LV

<400> 11
 attccccaga ccttctgcag attctgtggt tataactcact cctcatccca aaga atg 57
 Met
 -15
 15 aaa ttt acc act ctc ctc ttc ttg gca gct gta gca ggg gcc ctg gtc 105
 Lys Phe Thr Thr Leu Leu Phe Leu Ala Ala Val Ala Gly Ala Leu Val
 -10 -5 1
 20 tat gct gaa gat gcc tcc tct gac tcg acg ggt gct gat cct gcc cag 153
 Tyr Ala Glu Asp Ala Ser Ser Asp Ser Thr Gly Ala Asp Pro Ala Gln
 5 10 15
 gaa gct ggg acc tct aag cct aat gaa gag atc tca ggt cca gca gaa 201
 Glu Ala Gly Thr Ser Lys Pro Asn Glu Glu Ile Ser Gly Pro Ala Glu
 20 25 30
 25 cca gct tca ccc cca gag aca acc aca aca gcc cag gag act tcg gcg 249
 Pro Ala Ser Pro Pro Glu Thr Thr Thr Thr Ala Gln Glu Thr Ser Ala
 35 40 45 50
 gca gca gtt cag ggg aca gcc aag gtc acc tca agc agg cag gaa cta 297
 Ala Ala Val Gln Gly Thr Ala Lys Val Thr Ser Ser Arg Gln Glu Leu
 30 55 60 65
 aac ccc ctg aaa tcc ata gtg gag aaa agt atc tta cta aca gaa caa 345
 Asn Pro Leu Lys Ser Ile Val Glu Lys Ser Ile Leu Leu Thr Glu Gln
 70 75 80
 gcc ctt gca aaa gca gga aaa gga atg cac gga ggc gtg cca ggt gga 393
 35 Ala Leu Ala Lys Ala Gly Lys Gly Met His Gly Gly Val Pro Gly Gly
 85 90 95
 aaa caa ttc atc gaa aat gga agt gaa ttt gca caa aaa tta ctg aag 441
 Lys Gln Phe Ile Glu Asn Gly Ser Glu Phe Ala Gln Lys Leu Leu Lys
 100 105 110
 40 aaa ttc agt cta tta aaa cca tgg gca tgagaagctg aataatggga 488
 Lys Phe Ser Leu Leu Lys Pro Trp Ala
 115 120
 tcattggact taaagcctta aatacccttg tagcccagag ctattaaaac gaaagcatcc 548
 aaaaaaaaaa aaaaaaa 565

45 <210> 12
 <211> 1663
 <212> DNA
 <213> Homo sapiens

50 <220>
 <221> CDS
 <222> 152..475

55 <220>
 <221> sig_peptide
 <222> 152..244
 <223> Von Heijne matrix
 score 10.0910253445132
 60 seq LVLLLVTRSPVNA/CL

<400> 12
 atgtgtctgc tgccgccatt gtgcggcgct ggtccctca gagggttcct gctgctgccg 60


```

gtgccttgga ccctccccct cgtttctcgt tctactgccc caggagcccg gcgggtccgg 120
gactcccgtc cgtgccggtg cgggagcccg c atg tgg ctg tgg gag gac cag 172
                               Met Trp Leu Trp Glu Asp Gln
                               -30                               -25
5  ggc ggc ctc ctg ggc cct ttc tcc ttc ctg ctg cta gtg ctg ctg ctg 220
   Gly Gly Leu Leu Gly Pro Phe Ser Phe Leu Leu Leu Val Leu Leu Leu
                               -20                               -15                               -10
   gtg acg cgg agc ccg gtc aat gcc tgc ctc ctc acc ggc agc ctc ttc 268
   Val Thr Arg Ser Pro Val Asn Ala Cys Leu Leu Thr Gly Ser Leu Phe
10                               -5                               1                               5
   gtt cta ctg cgc gtc ttc agc ttt gag ccg gtg ccc tct tgc agg gcc 316
   Val Leu Leu Arg Val Phe Ser Phe Glu Pro Val Pro Ser Cys Arg Ala
                               10                               15                               20
   ctg cag gtg ctc aag ccc cgg gac cgc att tct gcc atc gcc cac cgt 364
15  Leu Gln Val Leu Lys Pro Arg Asp Arg Ile Ser Ala Ile Ala His Arg
   25                               30                               35                               40
   ggc ggc agc aam sag gcg ccc gag aac acg ctg gcg gcc att cgg cag 412
   Gly Gly Ser Xaa Xaa Ala Pro Glu Asn Thr Leu Ala Ala Ile Arg Gln
                               45                               50                               55
20  cta aga atg gag caa cag gcg tgg agt tgg aca ttg agt tta ctt ctg 460
   Leu Arg Met Glu Gln Gln Ala Trp Ser Trp Thr Leu Ser Leu Leu Leu
                               60                               65                               70
   acg gga ttc ctg tct taatgcacga taacacagta gataggacga ctgatgggac 515
   Thr Gly Phe Leu Ser
25                               75
   tgggcgattg tgtgatttga catttgaaca aattaggaag ctgaatcctg cagcaaacca 575
   cagactcagg aatgatttcc ctgatgaaaa gatccctacc ctaagggaag ctgttgcaga 635
   gtgcctaaac cataacctca caatcttctt tgatgtcaaa ggccatgcac acaaggctac 695
   tgaggctcta aagaaaatgt atatggaatt tcctcaactg tataataata gtgtggtctg 755
30  ttctttcttg ccagaagtta tctacaaggt aacattcggg atttttcttg tacatattag 815
   atgagacaaa cagatcgggg tgtaataaca gcattaactc acagaccttg gagcctaagc 875
   catacaggag atgggaaacc acgctatgat actttcttga aacattttat atttgttatg 935
   atggacattt tgctcgattg gagcatgcat aatatcttgt ggtacctgtg tgggaatttca 995
   gctttcctca tgcaaaagga ttttgtatcc ccggcctact tgaagaagtg gtcagctaaa 1055
35  ggaatccagg ttgttggttg gactgttaat acctttgatg aaaagagtta ctacgaatcc 1115
   catcttggtt ccagctatat cactgacagc atggtagaag actgcgaacc tcacttctag 1175
   actttcacgg tgggacgaaa cgggttcaga aactgccagg ggccctcatc agggatatca 1235
   aaataccctt tgtgctagcc caggccctgg ggaatcaggt gactcacaca aatgcaatag 1295
   ttggtcactg catttttacc tgaaccaaag ctaaaccagg tgttgccacc atgcaccatg 1355
40  gcatgccaga gttcaacact gttgctcttg aaaatctggg tctgaaaaaa cgcacaagag 1415
   cccctgccct gccctagctg aggcacacag ggagacccag tgaggataag cacagattga 1475
   attgtacaat ttgcagatgc agatgtaaat gcatgggaca tgcatgataa ctcagagttg 1535
   acattttaaa acttgccaca cttatttcaa atatttgtac tcagctatgt taacatgtac 1595
   tgtagacatc aaacttgtgg ccatactaata aaaattatta aaaggagcac taaaaaaaaa 1655
45  aaaaaaaaaa 1663

<210> 13
<211> 744
<212> DNA
50 <213> Homo sapiens

<220>
<221> CDS
<222> 112..552
55 <220>
   <221> sig_peptide
   <222> 112..183
   <223> Von Heijne matrix
60   score 11.7298925418815
   seq FVLGLGLTPPTLA/QD

<400> 13

```

```

tcacaactgg aacccatctc caggaacaaa cagctggaac ccatctcccg ttgaagggaa      60
actgccagat ttttgtaaga ttcttctctc tgggagcctg tgttggaaga g atg gtg      117
                                     Met Val
5  atg ggc ctg ggc gtt ttg ttg ttg gtc ttc gtg ctg ggt ctg ggt ctg      165
   Met Gly Leu Gly Val Leu Leu Leu Val Phe Val Leu Gly Leu Gly Leu
   -20 -15 -10
   acc cca ccg acc ctg gct cag gat aac tcc agg tac aca cac ttc ctg      213
   Thr Pro Pro Thr Leu Ala Gln Asp Asn Ser Arg Tyr Thr His Phe Leu
   -5 1 5 10
10 acc cag cac tat gat gcc aaa cca cag ggc cgg gat gac aga tac tgt      261
   Thr Gln His Tyr Asp Ala Lys Pro Gln Gly Arg Asp Asp Arg Tyr Cys
   15 20 25
   gaa agc atc atg agg aga cgg ggc ctg acc tca ccc tgc aaa gac atc      309
   Glu Ser Ile Met Arg Arg Arg Gly Leu Thr Ser Pro Cys Lys Asp Ile
   30 35 40
15 aac aca ttt att cat ggc aac aag cgc acg atc aag gcc atc tgt gaa      357
   Asn Thr Phe Ile His Gly Asn Lys Arg Thr Ile Lys Ala Ile Cys Glu
   45 50 55
   aac aag aat gga aac cct cac aga gaa aac cta aga ata agc aag tct      405
   Asn Lys Asn Gly Asn Pro His Arg Glu Asn Leu Arg Ile Ser Lys Ser
   60 65 70
   tct ttc cag gtc acc act tgc aag cta cat gga ggt tcc ccc tgg cct      453
   Ser Phe Gln Val Thr Thr Cys Lys Leu His Gly Gly Ser Pro Trp Pro
   75 80 85 90
25 cca tgc cag tac cga gcc aca gcg ggg ttc aga aac gtt gtt gtt gct      501
   Pro Cys Gln Tyr Arg Ala Thr Ala Gly Phe Arg Asn Val Val Val Ala
   95 100 105
   tgt gaa aat ggc tta cct gtc cac ttg gat cag tca att ttc cgt cgt      549
   Cys Glu Asn Gly Leu Pro Val His Leu Asp Gln Ser Ile Phe Arg Arg
   110 115 120
30 ccg taaccagcgg gcccttggtc aagtgtgtgc tctgtgtgtc ttgccttcca      602
   Pro
   tttccctctt gcacccagaa cagtgggtggc aacattcatt gccaaagggcc caaagaaaga      662
   gctacctgga ccttttgttt tctgtttgac aacatgttta ataaataaaa atgtcttgat      722
35 atcagcaaaa aaaaaaaaaa aa      744

```

<210> 14
 <211> 1759
 <212> DNA
 40 <213> Homo sapiens

<220>
 <221> CDS
 <222> 101..1243

45 <220>
 <221> sig_peptide
 <222> 101..199
 <223> Von Heijne matrix
 50 score 3.57142340200611
 seq FLCLGMALCPRQA/TR

```

<400> 14
gtagagtgtc gaaggctctg ccaacggctc tcttggcgtc tcaacgttcg gatcagcagc      60
55 ttttttccat tctctctctc cacttcttca gtgagcagcc atg agt tgg act gtg      115
                                     Met Ser Trp Thr Val
                                     -30
   cct gtt gtg cgg gcc agc cag aga gtg agc tcg gtg gga gcg aat ttc      163
   Pro Val Val Arg Ala Ser Gln Arg Val Ser Ser Val Gly Ala Asn Phe
   -25 -20 -15
   cta tgc ctg ggg atg gcc ctg tgt ccg cgt caa gca acg cgc atc ccg      211
   Leu Cys Leu Gly Met Ala Leu Cys Pro Arg Gln Ala Thr Arg Ile Pro
   -10 -5 1

```

	ctc	aac	ggc	acc	tgg	ctc	ttc	acc	ccc	gtg	agc	aag	atg	gcg	act	gtg	259
	Leu	Asn	Gly	Thr	Trp	Leu	Phe	Thr	Pro	Val	Ser	Lys	Met	Ala	Thr	Val	
	5					10					15				20		
	aag	agt	gag	ctt	att	gag	cgt	ttc	act	tcc	gag	aag	ccc	gtt	cat	cac	307
5	Lys	Ser	Glu	Leu	Ile	Glu	Arg	Phe	Thr	Ser	Glu	Lys	Pro	Val	His	His	
					25					30					35		
	agt	aag	gtc	tcc	atc	ata	gga	act	gga	tcg	gtg	ggc	atg	gcc	tgc	gct	355
	Ser	Lys	Val	Ser	Ile	Ile	Gly	Thr	Gly	Ser	Val	Gly	Met	Ala	Cys	Ala	
				40					45					50			
10	atc	agc	atc	tta	tta	aaa	ggc	ttg	agt	gat	gaa	ctt	gcc	ctt	gtg	gat	403
	Ile	Ser	Ile	Leu	Leu	Lys	Gly	Leu	Ser	Asp	Glu	Leu	Ala	Leu	Val	Asp	
				55				60					65				
	ctt	gat	gaa	gac	aaa	ctg	aag	ggg	gag	acg	atg	gat	ctt	caa	cat	ggc	451
	Leu	Asp	Glu	Asp	Lys	Leu	Lys	Gly	Glu	Thr	Met	Asp	Leu	Gln	His	Gly	
15		70				75					80						
	agc	cct	ttc	acg	aaa	atg	cca	aat	att	gtt	tgt	agc	aaa	gat	tac	ttt	499
	Ser	Pro	Phe	Thr	Lys	Met	Pro	Asn	Ile	Val	Cys	Ser	Lys	Asp	Tyr	Phe	
					90					95					100		
	gtc	aca	gca	aac	tcc	aac	cta	gtg	att	atc	aca	gca	ggg	gca	cgc	caa	547
20	Val	Thr	Ala	Asn	Ser	Asn	Leu	Val	Ile	Ile	Thr	Ala	Gly	Ala	Arg	Gln	
				105						110					115		
	gaa	aag	gga	gaa	acg	cgc	ctt	aat	tta	gtc	cag	cga	aat	gtg	gcc	atc	595
	Glu	Lys	Gly	Glu	Thr	Arg	Leu	Asn	Leu	Val	Gln	Arg	Asn	Val	Ala	Ile	
				120					125					130			
25	ttc	aag	tta	atg	att	tcc	agt	att	gtc	cag	tac	agc	ccc	cac	tgc	aaa	643
	Phe	Lys	Leu	Met	Ile	Ser	Ser	Ile	Val	Gln	Tyr	Ser	Pro	His	Cys	Lys	
				135				140					145				
	ctg	att	att	gtt	tcc	aat	cca	gtg	gat	atc	tta	act	tat	gta	gct	tgg	691
	Leu	Ile	Ile	Val	Ser	Asn	Pro	Val	Asp	Ile	Leu	Thr	Tyr	Val	Ala	Trp	
30		150				155					160						
	aag	ttg	agt	gca	ttt	ccc	aaa	aac	cgt	att	att	gga	agc	ggc	tgt	aat	739
	Lys	Leu	Ser	Ala	Phe	Pro	Lys	Asn	Arg	Ile	Ile	Gly	Ser	Gly	Cys	Asn	
		165			170				175						180		
	ctg	gat	act	gct	cgt	ttt	cgt	ttc	ttg	att	gga	caa	aag	ctt	ggg	atc	787
35	Leu	Asp	Thr	Ala	Arg	Phe	Arg	Phe	Leu	Ile	Gly	Gln	Lys	Leu	Gly	Ile	
				185					190						195		
	cat	tct	gaa	agc	tgc	cat	gga	tgg	atc	ctc	gga	gag	cat	gga	gac	tca	835
	His	Ser	Glu	Ser	Cys	His	Gly	Trp	Ile	Leu	Gly	Glu	His	Gly	Asp	Ser	
				200					205					210			
40	agt	gtt	cct	gtg	tgg	agt	gga	gtg	aac	ata	gct	ggg	gtc	cct	ttg	aag	883
	Ser	Val	Pro	Val	Trp	Ser	Gly	Val	Asn	Ile	Ala	Gly	Val	Pro	Leu	Lys	
				215				220						225			
	gat	ctg	aac	tct	gat	ata	gga	act	gat	aaa	gat	cct	gag	caa	tgg	aaa	931
	Asp	Leu	Asn	Ser	Asp	Ile	Gly	Thr	Asp	Lys	Asp	Pro	Glu	Gln	Trp	Lys	
45		230				235						240					
	aat	gtc	cac	aaa	gaa	gtg	act	gca	act	gcc	tat	gag	att	att	aaa	atg	979
	Asn	Val	His	Lys	Glu	Val	Thr	Ala	Thr	Ala	Tyr	Glu	Ile	Ile	Lys	Met	
		245			250					255					260		
	aaa	ggg	tat	act	tct	tgg	gcc	att	ggc	cta	tct	gtg	gcc	gat	tta	aca	1027
50	Lys	Gly	Tyr	Thr	Ser	Trp	Ala	Ile	Gly	Leu	Ser	Val	Ala	Asp	Leu	Thr	
				265					270						275		
	gaa	agt	att	ttg	aag	aat	ctt	agg	aga	ata	cat	cca	gtt	tcc	acc	ata	1075
	Glu	Ser	Ile	Leu	Lys	Asn	Leu	Arg	Arg	Ile	His	Pro	Val	Ser	Thr	Ile	
				280					285					290			
55	att	aag	ggc	ctc	tat	gga	ata	gat	gaa	gaa	gta	ttc	ctc	agt	att	cct	1123
	Ile	Lys	Gly	Leu	Tyr	Gly	Ile	Asp	Glu	Glu	Val	Phe	Leu	Ser	Ile	Pro	
				295			300						305				
	tgt	atc	ctg	gga	gag	aac	ggg	att	acc	aac	ctt	ata	aag	ata	aag	ctg	1171
	Cys	Ile	Leu	Gly	Glu	Asn	Gly	Ile	Thr	Asn	Leu	Ile	Lys	Ile	Lys	Leu	
60		310				315						320					
	acc	cct	gaa	gaa	gag	gcc	cat	ctg	aaa	aaa	agt	gca	aaa	aca	ctc	tgg	1219
	Thr	Pro	Glu	Glu	Glu	Ala	His	Leu	Lys	Lys	Ser	Ala	Lys	Thr	Leu	Trp	
		325				330					335					340	

gaa att cag aat aag ctt aag ctt taaagttgcc' taaaactacc attccgaaat 1273
 Glu Ile Gln Asn Lys Leu Lys Leu
 345

5 tattgaagag atcatagata caggattata taacgaaatt ttgaataaac ttgaattcct 1333
 aaaagatgga aacaggaaag taggttagagt gatttttccta tttatttagt cctccagctc 1393
 ttttattgag catccacgtg ctggacgata cttattttaca attcctaagt atttttggtg 1453
 cctctgatgt agcagcactt gccatgttat atatatgtag ttggcatttg gttcccaaaa 1513
 agtaggatgt aggtatttat tgtgttctag aaattccgac tcttttcatt agatatatgc 1573
 tattttctttc attcttgctg gtttatacct atgttcattt atatgctgta aaaaagtagt 1633
 10 agcttcttct acaatgtaaa aataaatgta catacaaaaa aatgcagtag tatatacaat 1693
 cttttgtttt gcttcctttg atagttaata aattccgttt gttgaatcaa taaaaaaaaa 1753
 aaaaaa 1759

<210> 15
 15 <211> 1755
 <212> DNA
 <213> Homo sapiens

<220>
 20 <221> CDS
 <222> 101..517

<220>
 <221> sig_peptide
 25 <222> 101..199
 <223> Von Heijne matrix
 score 3.57613483592743
 seq FLCLGMALCLRQA/TR

30 <400> 15
 gtagagtgc tgaaggtcctg ccaacggctc tcttggcgtc tcaacgttcg gatcagcagc 60
 ttttttccat tctctctctc cacttcttca gtgagcagcc atg agt tgg act gtg 115
 Met Ser Trp Thr Val
 -30

35 cct gtt gtg cgg gcc agc cag aga atg agc tcg gtg gga gcg aat ttc 163
 Pro Val Val Arg Ala Ser Gln Arg Met Ser Ser Val Gly Ala Asn Phe
 -25 -20 -15

40 cta tgc ctg ggg atg gcc ctg tgt ctg cgt caa gca acg cgc atc ccg 211
 Leu Cys Leu Gly Met Ala Leu Cys Leu Arg Gln Ala Thr Arg Ile Pro
 -10 -5 1

ctc aac ggc acc tgg ctc ttc aca ccc gtg agc aag atg gcg act gtg 259
 Leu Asn Gly Thr Trp Leu Phe Thr Pro Val Ser Lys Met Ala Thr Val
 5 10 15 20

45 aag agt gag ctt att gag cgt ttc act tcc gag aag ccc gtt cat cac 307
 Lys Ser Glu Leu Ile Glu Arg Phe Thr Ser Glu Lys Pro Val His His
 25 30 35

agt aag gtc tcc atc ata gga act gga tcg gtg ggc atg gcc tgc gct 355
 Ser Lys Val Ser Ile Ile Gly Thr Gly Ser Val Gly Met Ala Cys Ala
 40 45 50

50 atc agc atc ttg tta aaa ggc ttg agt gat gaa ctt gcc ctt gtg gat 403
 Ile Ser Ile Leu Leu Lys Gly Leu Ser Asp Glu Leu Ala Leu Val Asp
 55 60 65

ctt gat gaa gac aaa ctg aag ggt gag acg atg gat ctt caa cat ggc 451
 Leu Asp Glu Asp Lys Leu Lys Gly Glu Thr Met Asp Leu Gln His Gly
 70 75 80

55 agc cct ttc acg aaa atg cca ata ttg ttt gta gca aag att act ttg 499
 Ser Pro Phe Thr Lys Met Pro Ile Leu Phe Val Ala Lys Ile Thr Leu
 85 90 95 100

tca cag caa act cca acc tagtgattat cacagcaggt gcacgccaaag 547
 60 Ser Gln Gln Thr Pro Thr
 105

aaaagggaga aacgcgcctt aatttagtcc agcgaaatgt ggccatcttc aagtaatgat 607
 ttccagtatt gtccagtaca gccccactg caaactgatt attgtttcca atccagtgga 667

```

tattcttaact tatgtagctt ggaagttgag tgcatttccc aaaaaccgta ttattggaag 727
cggctgtaat ctggatactg ctcgttttcg tttcttgatt ggacaaaagc ttggatatcca 787
ttctgaaagc tgccatggat ggatcctcgg agagcatgga gactcaagtg ttcctgtgtg 847
gagtgaggagtg aacatagctg gtgtcccttt gaaggatctg aactctgata taggaactga 907
5 taaagatcct gagcaggaaa aatgtccaca aagaagtgc tgcaactgcc tatgagatta 967
ttaaagtgaaggagggttatact tcttgggcca ttggcctatc tgtggccgat ttaacagaaa 1027
gtatttttgaa gaatccttagg agaatacatc cagtttccac cataactaag ggcctctatg 1087
gaatagatga agaagtattc ctcagtattc cttgtatcct gggagagaac ggtattacca 1147
accttataaaa gataaagctg acccctgaag aagaggccca tctgaaaaaa agtgcaaaaa 1207
10 cactctggga aattcagaat aagcttaagc tttaaagttg cctaaaaacta ccattccgaa 1267
attattgaag agatcataga tacaggatta tataacgaaa ttttgaataa acttgaattc 1327
ctaaaagatg gaaacaggaa agtaggtaga gtgattttcc tatttattta gtcctccagc 1387
tcttttattg agcatccacg tgctggacga tacttattta caattcctaa gtatttttgg 1447
tacctctgat gtagcagcac ttgccatgtt atatatatgt agttggcatt tggttcccaa 1507
15 aaagtaggat gtaggtattt attgtgttct agaaattccg actcttttca ttagatatat 1567
gctattttctt tcattcttgc tggtttatac ctatgttcat ttatatgctg taaaaaagta 1627
gtagcttctt ctacaatgta aaaataaatg tacatacaaa aaaatgcagt agtatataca 1687
atcttttgtt ttgcttcctt tgatagttaa taaattccgt ttgttgaatc aataaaaaaa 1747
aaaaaaaaa 1755

20
<210> 16
<211> 936
<212> DNA
<213> Homo sapiens

25
<220>
<221> CDS
<222> 59..853

30
<220>
<221> sig_peptide
<222> 59..100
<223> Von Heijne matrix
score 5.2402423806254

35
seq NFILFIFIPGVFS/LK

<400> 16
agaaaggagg ctctgggtag acgcactaga ttactggata aatcacttca atttccca 58
atg aat ttt ata ttg ttt att ttt ata cct gga gtt ttt tcc tta aaa 106
40 Met Asn Phe Ile Leu Phe Ile Phe Ile Pro Gly Val Phe Ser Leu Lys
-10 -5 1
agt agc act ttg aag cct act att gaa gca ttg cct aat gtg cta cct 154
Ser Ser Thr Leu Lys Pro Thr Ile Glu Ala Leu Pro Asn Val Leu Pro
5 10 15
45 tta aat gaa gat gtt aat aag cag gaa gaa aag aat gaa gat cat act 202
Leu Asn Glu Asp Val Asn Lys Gln Glu Glu Lys Asn Glu Asp His Thr
20 25 30
ccc aat tat gct cct gct aat gag aaa aat ggc aat tat tat aaa gat 250
Pro Asn Tyr Ala Pro Ala Asn Glu Lys Asn Gly Asn Tyr Tyr Lys Asp
50 35 40 45 50
ata aaa caa tat gtg ttc aca aca caa aat cca aat ggc act gag tct 298
Ile Lys Gln Tyr Val Phe Thr Thr Gln Asn Pro Asn Gly Thr Glu Ser
55 60 65
gaa ata tct gtg aga gcc aca act gac ctg aat ttt gct cta aaa aac 346
55 Glu Ile Ser Val Arg Ala Thr Thr Asp Leu Asn Phe Ala Leu Lys Asn
70 75 80
gga tca acc cca aac gtg cct gca ttt tgg aca atg tta gct aaa gct 394
Gly Ser Thr Pro Asn Val Pro Ala Phe Trp Thr Met Leu Ala Lys Ala
85 90 95
60 ata aat gga aca gca gtg gtc atg gat gat aaa gat caa tta ttt cac 442
Ile Asn Gly Thr Ala Val Val Met Asp Asp Lys Asp Gln Leu Phe His
100 105 110
cca att cca gag tct gat gtg aat gct aca cag gga gaa aat cag cca 490

```

	Pro	Ile	Pro	Glu	Ser	Asp	Val	Asn	Ala	Thr	Gln	Gly	Glu	Asn	Gln	Pro	
	115					120					125					130	
	gat	cta	gag	gat	ctg	aag	atc	aaa	ata	atg	ctg	gga	atc	tcg	ttg	atg	538
5	Asp	Leu	Glu	Asp	Leu	Lys	Ile	Lys	Ile	Met	Leu	Gly	Ile	Ser	Leu	Met	
					135					140					145		
	acc	ctc	ctc	ctc	ttt	gtg	gtc	ctc	ttg	gca	ttc	tgt	agt	gct	aca	ctg	586
	Thr	Leu	Leu	Leu	Phe	Val	Val	Leu	Leu	Ala	Phe	Cys	Ser	Ala	Thr	Leu	
					150					155					160		
	tac	aaa	ctg	agg	cat	ctg	agt	tat	aaa	agt	tgt	gag	agt	cag	tac	tct	634
10	Tyr	Lys	Leu	Arg	His	Leu	Ser	Tyr	Lys	Ser	Cys	Glu	Ser	Gln	Tyr	Ser	
			165					170						175			
	gtc	aac	cca	gag	ctg	gcc	acg	atg	tct	tac	ttt	cat	cca	tca	gaa	ggc	682
	Val	Asn	Pro	Glu	Leu	Ala	Thr	Met	Ser	Tyr	Phe	His	Pro	Ser	Glu	Gly	
		180					185					190					
15	ggt	tca	gat	aca	tcc	ttt	tcc	aag	agt	gca	gag	agc	agc	aca	ttt	ttg	730
	Val	Ser	Asp	Thr	Ser	Phe	Ser	Lys	Ser	Ala	Glu	Ser	Ser	Thr	Phe	Leu	
						200					205					210	
	ggc	acc	act	tct	tca	gat	atg	aga	aga	tca	ggc	aca	aga	aca	tca	gaa	778
	Gly	Thr	Thr	Ser	Ser	Asp	Met	Arg	Arg	Ser	Gly	Thr	Arg	Thr	Ser	Glu	
20						215					220					225	
	tct	aag	ata	atg	acg	gat	atc	att	tcc	ata	ggc	tca	gat	aat	gag	atg	826
	Ser	Lys	Ile	Met	Thr	Asp	Ile	Ile	Ser	Ile	Gly	Ser	Asp	Asn	Glu	Met	
					230				235					240			
25	cat	gaa	aac	gat	gag	tcg	gtt	acc	cgg	tgaagaaatc	aaggaacccg						873
	His	Glu	Asn	Asp	Glu	Ser	Val	Thr	Arg								
			245					250									
	gtgaagaaat	cttattgatg	aataaataac	tttaattatt	ttgtcatcaa	aaaaaaaaaa											933
	aaa																936
30	<210> 17																
	<211> 747																
	<212> DNA																
	<213> Homo sapiens																
35	<220>																
	<221> CDS																
	<222> 73..672																
	<220>																
40	<221> sig_peptide																
	<222> 73..132																
	<223> Von Heijne matrix																
	score 5.21332530399231																
	seq SPVFLVFPPEITA/SE																
45	<400> 17																
	acaagaaaag	aacatggtct	agactgaagt	accaactaaa	tcattctcctt	tcaaattatc											60
	accgacacca	tc	atg	gat	tca	agc	acc	gca	cac	agt	ccg	gtg	ttt	ctg	gta		111
			Met	Asp	Ser	Ser	Thr	Ala	His	Ser	Pro	Val	Phe	Leu	Val		
50			-20					-15						-10			
	ttt	cct	cca	gaa	atc	act	gct	tca	gaa	tat	gag	tcc	aca	gaa	ctt	tca	159
	Phe	Pro	Pro	Glu	Ile	Thr	Ala	Ser	Glu	Tyr	Glu	Ser	Thr	Glu	Leu	Ser	
			-5				1				5						
	gcc	acg	acc	ttt	tca	act	caa	agc	ccc	ttg	caa	aaa	tta	ttt	gct	aga	207
55	Ala	Thr	Thr	Phe	Ser	Thr	Gln	Ser	Pro	Leu	Gln	Lys	Leu	Phe	Ala	Arg	
	10					15				20					25		
	aaa	atg	aaa	atc	tta	ggg	act	atc	cag	atc	ctg	ttt	gga	att	atg	acc	255
	Lys	Met	Lys	Ile	Leu	Gly	Thr	Ile	Gln	Ile	Leu	Phe	Gly	Ile	Met	Thr	
					30				35					40			
60	ttt	tct	ttt	gga	gtt	atc	ttc	ctt	ttc	act	ttg	tta	aaa	cca	tat	cca	303
	Phe	Ser	Phe	Gly	Val	Ile	Phe	Leu	Phe	Thr	Leu	Leu	Lys	Pro	Tyr	Pro	
				45				50					55				
	agg	ttt	ccc	ttt	ata	ttt	ctt	tca	gga	tat	cca	ttc	tgg	ggc	tct	gtt	351

	Arg	Phe	Pro	Phe	Ile	Phe	Leu	Ser	Gly	Tyr	Pro	Phe	Trp	Gly	Ser	Val	
			60					65					70				
	ttg	ttc	att	aat	tct	gga	gcc	ttc	cta	att	gca	gtg	aaa	aga	aaa	acc	399
	Leu	Phe	Ile	Asn	Ser	Gly	Ala	Phe	Leu	Ile	Ala	Val	Lys	Arg	Lys	Thr	
5		75					80					85					
	aca	gaa	act	ctg	ata	ata	ttg	agc	cga	ata	atg	aat	ttt	ctt	agt	gcc	447
	Thr	Glu	Thr	Leu	Ile	Ile	Leu	Ser	Arg	Ile	Met	Asn	Phe	Leu	Ser	Ala	
	90					95					100				105		
	ctg	gga	gca	ata	gct	gga	atc	att	ctc	ctc	aca	ttt	ggg	ttc	atc	cta	495
10	Leu	Gly	Ala	Ile	Ala	Gly	Ile	Ile	Leu	Leu	Thr	Phe	Gly	Phe	Ile	Leu	
					110					115					120		
	gat	caa	aac	tac	att	tgt	ggg	tat	tct	cac	caa	aat	agt	cag	tgt	aag	543
	Asp	Gln	Asn	Tyr	Ile	Cys	Gly	Tyr	Ser	His	Gln	Asn	Ser	Gln	Cys	Lys	
				125					130					135			
15	gct	gtt	act	gtc	ctg	ttc	ttg	gga	att	ttg	att	aca	ttg	atg	act	ttc	591
	Ala	Val	Thr	Val	Leu	Phe	Leu	Gly	Ile	Leu	Ile	Thr	Leu	Met	Thr	Phe	
				140				145					150				
	agc	att	att	gaa	tta	ttc	att	tct	ctg	cct	ttc	tca	att	ttg	ggg	tgc	639
	Ser	Ile	Ile	Glu	Leu	Phe	Ile	Ser	Leu	Pro	Phe	Ser	Ile	Leu	Gly	Cys	
20		155					160					165					
	cac	tca	gag	gat	tgt	gat	tgt	gaa	caa	tgt	tgt	tgactagcac	tggtgagaata				692
	His	Ser	Glu	Asp	Cys	Asp	Cys	Glu	Gln	Cys	Cys						
	170					175					180						
	aagatgtgtt	aaaatatttaa	aaaaaaaaaaaa	aaaaaaaaaaag	aaaaaaaaaaaa	aaaaaa											747
25	<210>	18															
	<211>	1884															
	<212>	DNA															
	<213>	Homo sapiens															
30	<220>																
	<221>	CDS															
	<222>	94..1275															
35	<220>																
	<221>	sig_peptide															
	<222>	94..210															
	<223>	Von Heijne matrix															
		score	4.55778392992629														
40		seq	LVLVKRLLAVSVS/CI														
	<400>	18															
	acagcgcgtg	cagcctcggtg	cagctcttct	gggtctccggc	gccccgccct	cagacgtaat											60
45	gttgaattaa	agaaaatact	ttatcagaag	aag	atg	gcc	act	gcc	cag	ttg	cag						114
			Met	Ala	Thr	Ala	Gln	Leu	Gln								
	agg	act	ccc	atg	agt	gca	ctg	gta	ttt	ccc	aat	aag	ata	tca	act	gaa	162
	Arg	Thr	Pro	Met	Ser	Ala	Leu	Val	Phe	Pro	Asn	Lys	Ile	Ser	Thr	Glu	
			-30					-25					-20				
50	cac	cag	tct	ttg	gtg	tta	gtg	aag	agg	ctt	cta	gca	gtt	tca	gta	tcc	210
	His	Gln	Ser	Leu	Val	Leu	Val	Lys	Arg	Leu	Leu	Ala	Val	Ser	Val	Ser	
			-15				-10					-5					
	tgt	atc	acg	tat	ttg	agg	gga	ata	ttc	cca	gaa	tgc	gct	tat	gga	aca	258
	Cys	Ile	Thr	Tyr	Leu	Arg	Gly	Ile	Phe	Pro	Glu	Cys	Ala	Tyr	Gly	Thr	
55	1			5						10					15		
	aga	tat	cta	gat	gat	ctt	tgt	gtc	aaa	ata	ctg	aga	gaa	gat	aaa	aat	306
	Arg	Tyr	Leu	Asp	Asp	Leu	Cys	Val	Lys	Ile	Leu	Arg	Glu	Asp	Lys	Asn	
				20					25					30			
	tgc	cca	gga	tct	aca	cag	tta	gtg	aaa	tgg	att	cta	gga	tgt	tat	gat	354
60	Cys	Pro	Gly	Ser	Thr	Gln	Leu	Val	Lys	Trp	Ile	Leu	Gly	Cys	Tyr	Asp	
			35					40					45				
	gct	tta	cag	aaa	aaa	tat	cta	agg	atg	ggt	ggt	cta	gct	gta	tac	aca	402
	Ala	Leu	Gln	Lys	Lys	Tyr	Leu	Arg	Met	Val	Val	Leu	Ala	Val	Tyr	Thr	

23


```

ttggtattaa aatttggtat tgattcactt tttacttatg ttaaaattat accatttaac 1675
tggctctttt gtcattgtgc tgttattaaa acaatgttct tcaatatttt gacataatgt 1735
attaacattt taatatataa tgtacaattt aagaattggt gctttacctt tactatgctt 1795
tttttacagg acaaaaagac tgatttttaa agtatggcat tttttgcagc ataaataaaa 1855
5 tattgttcag tacgaaaaaa aaaaaaaaaa 1884

<210> 19
<211> 691
<212> DNA
10 <213> Homo sapiens

<220>
<221> CDS
<222> 42..515
15
<220>
<221> sig_peptide
<222> 42..92
<223> Von Heijne matrix
20 score 10.7019149919754
seq VLMLLAVLIWTGA/EN

<400> 19
25 gagttgtcct gtgctggagg tctgctcaga cgaagggtctc c atg gcg tta gaa gtc 56
Met Ala Leu Glu Val
-15
ttg atg ctc ctc gct gtc ttg att tgg acc ggt gct gag aac ctc cat 104
Leu Met Leu Leu Ala Val Leu Ile Trp Thr Gly Ala Glu Asn Leu His
-10 -5 1
30 gtg aaa ata agt tgc tct ctg gac tgg ttg atg gtc tca gtt atc cca 152
Val Lys Ile Ser Cys Ser Leu Asp Trp Leu Met Val Ser Val Ile Pro
5 10 15 20
gtt gca gaa agc aga aat ctg tat ata ttt gcg gat gaa tta cat ctg 200
Val Ala Glu Ser Arg Asn Leu Tyr Ile Phe Ala Asp Glu Leu His Leu
35 25 30 35
gga atg ggc tgc cct gca aat cgg ata cat aca tat gta tat gag ttt 248
Gly Met Gly Cys Pro Ala Asn Arg Ile His Thr Tyr Val Tyr Glu Phe
40 40 45 50
ata tat ctt gtt cgt gat tgt ggc atc agg aca agg gta gtt tct gag 296
Ile Tyr Leu Val Arg Asp Cys Gly Ile Arg Thr Arg Val Val Ser Glu
55 60 65
gaa act ctc ctt ttt caa acc gag ctg tac ttt acc cca agg aat ata 344
Glu Thr Leu Leu Phe Gln Thr Glu Leu Tyr Phe Thr Pro Arg Asn Ile
70 75 80
45 gat cat gac cct cag gaa atc cat ttg gag tgt tcc acc tct agg aaa 392
Asp His Asp Pro Gln Glu Ile His Leu Glu Cys Ser Thr Ser Arg Lys
85 90 95 100
tca gtg tgg ctt aca cca gtt tct act gag aat gaa ata aaa ttg gat 440
Ser Val Trp Leu Thr Pro Val Ser Thr Glu Asn Glu Ile Lys Leu Asp
105 110 115
50 cct agt cct ttt att gct gac ttt cag aca aca gca gaa gag tta gga 488
Pro Ser Pro Phe Ile Ala Asp Phe Gln Thr Thr Ala Glu Glu Leu Gly
120 125 130
tta tta tct tct agt cca aac ttg ctc tgagctaaag gagaaatgga 535
55 Leu Leu Ser Ser Pro Asn Leu Leu
135 140
aacttgaagc tgggtgttatg tatttttcag gaaaacagtt tcatttttttc atagcaaaaa 595
tatagttggg gtatatctct ccttaagtct ctggtttcta aaaaccctac ttcagtaaaag 655
gtcctgatta gttgattagc gaaaaaaaaa aaaaaa 691
60
<210> 20
<211> 1138
<212> DNA

```

<213> Homo sapiens

<220>

<221> CDS

5 <222> 271..969

<220>

<221> sig_peptide

<222> 271..366

10 <223> Von Heijne matrix
 score 5.6680378526706
 seq WMGLACFRSLAAS/SP

<220>

15 <221> misc_feature

<222> 989

<223> n=a, g, c or t

<400> 20

20	aaaaaccttt caagtgcgcc ctccttttct taaagtcttt tataggggtc cccttcttgg	60
	ccatctccat cctgtgagtc aggactgaaa gggcacagac aggtcactgc cagcattgtt	120
	ggggcaagcc tgcaagcag catcactggg gatctgacat gacaatggcc gcctgcccc	180
	tctgagggct acaggactta cccagtgagg aagcagctaa gcaggtctga ccagccgacc	240
25	tggacctggc caagggtcct gtcacccctc atg gcc acc ccg cca ttc cgg ctg	294
	Met Ala Thr Pro Pro Phe Arg Leu	
	-30 -25	
	ata agg aag atg ttt tcc ttc aag gtg agc aga tgg atg ggg ctt gcc	342
	Ile Arg Lys Met Phe Ser Phe Lys Val Ser Arg Trp Met Gly Leu Ala	
	-20 -15 -10	
30	tgc ttc cgg tcc ctg gcg gca tcc tct ccc agt att cgc cag aag aaa	390
	Cys Phe Arg Ser Leu Ala Ala Ser Ser Pro Ser Ile Arg Gln Lys Lys	
	-5 1 5	
	cta atg cac aag ctg cag gag gaa aag gct ttt cgc gaa gag atg aaa	438
	Leu Met His Lys Leu Gln Glu Glu Lys Ala Phe Arg Glu Glu Met Lys	
35	10 15 20	
	att ttt cgt gaa aaa ata gag gac ttc agg gaa gag atg tgg act ttc	486
	Ile Phe Arg Glu Lys Ile Glu Asp Phe Arg Glu Glu Met Trp Thr Phe	
	25 30 35 40	
	cga ggc aag atc cat gct ttc cgg ggc cag atc ctg ggt ttt tgg gaa	534
40	Arg Gly Lys Ile His Ala Phe Arg Gly Gln Ile Leu Gly Phe Trp Glu	
	45 50 55	
	gag gag aga cct ttc tgg gaa gag gag aaa acc ttc tgg aaa gag gaa	582
	Glu Glu Arg Pro Phe Trp Glu Glu Glu Lys Thr Phe Trp Lys Glu Glu	
	60 65 70	
45	aaa tcc ttc tgg gaa atg gaa aag tct ttc agg gag gaa gag aaa act	630
	Lys Ser Phe Trp Glu Met Glu Lys Ser Phe Arg Glu Glu Lys Thr	
	75 80 85	
	ttc tgg aaa aag tac cgc act ttc tgg aag gag gat aag gcc ttc tgg	678
	Phe Trp Lys Lys Tyr Arg Thr Phe Trp Lys Glu Asp Lys Ala Phe Trp	
50	90 95 100	
	aaa gag gac aat gcc tta tgg gaa aga gac cgg aac ctt ctt cag gag	726
	Lys Glu Asp Asn Ala Leu Trp Glu Arg Asp Arg Asn Leu Leu Gln Glu	
	105 110 115 120	
	gac aag gcc ctg tgg gag gaa gaa aag gcc ctg tgg gta gag gaa aga	774
55	Asp Lys Ala Leu Trp Glu Glu Glu Lys Ala Leu Trp Val Glu Glu Arg	
	125 130 135	
	gcc ctc ctt gag ggg gag aaa gcc ctg tgg gaa gat aaa acg tcc ctc	822
	Ala Leu Leu Glu Gly Glu Lys Ala Leu Trp Glu Asp Lys Thr Ser Leu	
	140 145 150	
60	tgg gag gaa gag aat gcc ctc tgg gag gaa gag agg gcc ttc tgg atg	870
	Trp Glu Glu Glu Asn Ala Leu Trp Glu Glu Glu Arg Ala Phe Trp Met	
	155 160 165	
	gag aac aat ggc cac att gcc gga gag cag atg ctc gaa gat ggg ccc	918

Glu Asn Asn Gly His Ile Ala Gly Glu Gln Met Leu Glu Asp Gly Pro
 170 175 180
 cac aac gcc aac aga ggg cag cgc ttg ctg gcc ttc tcc cga ggc agg 966
 His Asn Ala Asn Arg Gly Gln Arg Leu Leu Ala Phe Ser Arg Gly Arg
 5 185 190 195 200
 gcg tagccagcat gcaggtgcan gggccctgtg gtccagactc ccctggggtg 1019
 Ala
 ggattcaagt ccaggggtgag cccatgtgct ggagaaaata cacactcatt ggtctccttg 1079
 ctttgaaaga tccaataaag tcctgaggca aggtttggaa aaccaaaaaa aaaaaaaaaa 1138
 10
 <210> 21
 <211> 468
 <212> DNA
 <213> Homo sapiens
 15
 <220>
 <221> CDS
 <222> 76..276
 20 <220>
 <221> sig_peptide
 <222> 76..135
 <223> Von Heijne matrix
 score 5.21332530399231
 25 seq SPVFLVFPPEITA/SE
 <400> 21
 agcacaagaa aagaacatgg tctagactga agtaccact aaatcatctc ctttcaaatt 60
 atcaccgaca ccatc atg gat tca agc acc gca cac agt ccg gtg ttt ctg 111
 30 Met Asp Ser Ser Thr Ala His Ser Pro Val Phe Leu
 -20 -15 -10
 gta ttt cct cca gaa atc act gct tca gaa tat gag tcc aca gaa ctt 159
 Val Phe Pro Pro Glu Ile Thr Ala Ser Glu Tyr Glu Ser Thr Glu Leu
 -5 1 5
 35 tca gcc acg acc ttt tca act caa agc ccc ttg caa aaa tta ttt gct 207
 Ser Ala Thr Thr Phe Ser Thr Gln Ser Pro Leu Gln Lys Leu Phe Ala
 10 15 20
 aga aaa atg aaa atc tta ggg gat atc cat tct ggg gct ctg ttt tgt 255
 Arg Lys Met Lys Ile Leu Gly Asp Ile His Ser Gly Ala Leu Phe Cys
 40 25 30 35 40
 tca tta att ctg gag cct tcc taattgcagt gaaaagaaaa accacagaaa 306
 Ser Leu Ile Leu Glu Pro Ser
 45
 ctctgggaat tttgattaca ttgatgactt tcagcattat tgaattattc atttctctgc 366
 45 ctttctcaat tttggggtgc cactcagagg attgtgattg tgaacaatgt tggtgactag 426
 cactgtgaga ataaagatgt gttaaaataa aaaaaaaaaa aa 468
 <210> 22
 <211> 720
 50 <212> DNA
 <213> Homo sapiens
 <220>
 <221> CDS
 55 <222> 6..287
 <220>
 <221> sig_peptide
 <222> 6..80
 60 <223> Von Heijne matrix
 score 4.17710408129886
 seq ISLSHLFLDLSRS/LW

```

<400> 22
atttg atg tgc ttc tta gtc tgc ttt aac ttg ccg att cat ata tcc ctg      50
Met Cys Phe Leu Val Ser Phe Asn Leu Pro Ile His Ile Ser Leu
-25 -20 -15
5 tct cat ttg ttc tta gat ttg tca cga agc ctc tgg ttt ttg gct tgt      98
Ser His Leu Phe Leu Asp Leu Ser Arg Ser Leu Trp Phe Leu Ala Cys
-10 -5 1 5
cct ggt ttg aac ttg gtg tat ctg gct ctt gac tca ttt tct gac ctc      146
Pro Gly Leu Asn Leu Val Tyr Leu Ala Leu Asp Ser Phe Ser Asp Leu
10 10 15 20
aga cca tcc tta aat ctg ctt ttc tac ttt gta cca ggc ttt ggc gtc      194
Arg Pro Ser Leu Asn Leu Leu Phe Tyr Phe Val Pro Gly Phe Gly Val
25 30 35
tcc aag tac ctg acc tca gct caa cct gtc ttg ggt ttt ctt ctc ctc      242
15 Ser Lys Tyr Leu Thr Ser Ala Gln Pro Val Leu Gly Phe Leu Leu Leu
40 45 50
cct gac att gac aac cca gcc ctc cta ggc aca gag aga tgg agc      287
Pro Asp Ile Asp Asn Pro Ala Leu Leu Gly Thr Glu Arg Trp Ser
55 60 65
20 tgaagtgtggt tttcctgaaa taaagcttgc attatgagag ggaataaaca gaagaaaaaa      347
atagtaagta aaatcttgct tgcctctcag taaaataaag ctctatTTTT cgtTTTTTTT      407
ttttccaact tcctgtacaa aaaagggaaa actttagctt ttgggggaaa tttggagcta      467
gcctgttggt actgttgagc ttagtgtatc tataactata tattattcca caatatctta      527
aatactttat aaagatatTT tcataaatta cagcaatcct ggcttttagat gattgatggc      587
25 catttttaaa caattaaagc taatttctag ctttttatga gtttggtatt aagcacagta      647
gtttcttaga aagtctccag ggaatgcatt ttgcaaaata aaaatcagct aatgacccaa      707
aaaaaaaaaaa aaa      720

<210> 23
30 <211> 727
<212> DNA
<213> Homo sapiens

<220>
35 <221> CDS
<222> 171..692

<220>
<221> sig_peptide
40 <222> 171..227
<223> Von Heijne matrix
score 4.17573075349936
seq LLLGQRCSLKVSG/QE

45 <400> 23
attgtgacat caccgtgcac tagccaatgg ctgcctgcct aagctgggtc cctgggtctcc      60
tgggactact agccctttgt tgatagggag aagccaacat ctcccgcagg accccctaatt      120
cttcagggca gctcccagag catggatccc tcctgattcc actcagcccg atg ttc      176
Met Phe
50 ctc aca gtc aag ctg ctc ctg ggc cag aga tgc agt ctg aag gtg tca      224
Leu Thr Val Lys Leu Leu Leu Gly Gln Arg Cys Ser Leu Lys Val Ser
-15 -10 -5
ggg caa gag agt gta gcc acg ctg aag aga ctg gtg tcc agg cgg ctg      272
Gly Gln Glu Ser Val Ala Thr Leu Lys Arg Leu Val Ser Arg Arg Leu
55 1 5 10 15
aag gtg cct gag gag cag cag cac ctg ctt ttc cgt ggc cag ctc ctg      320
Lys Val Pro Glu Glu Gln Gln His Leu Leu Phe Arg Gly Gln Leu Leu
20 25 30
gag gat gac aag cac ctc tct gac tac tgc att ggg ccc aat gcc tct      368
60 Glu Asp Asp Lys His Leu Ser Asp Tyr Cys Ile Gly Pro Asn Ala Ser
35 40 45
atc aat gtc atc atg cag ccc ttg gag aag atg gcg cta aag gag gcc      416
Ile Asn Val Ile Met Gln Pro Leu Glu Lys Met Ala Leu Lys Glu Ala

```

```

      50      55      60
cac cag ccg cag acc cag ccc ctg tgg cac cag ctg gga ctg gtc cta 464
His Gln Pro Gln Thr Gln Pro Leu Trp His Gln Leu Gly Leu Val Leu
      65      70      75
5 gct aaa cac ttt gaa cca cag gat gcc aag gcc gtg ctg cag ctg cta 512
Ala Lys His Phe Glu Pro Gln Asp Ala Lys Ala Val Leu Gln Leu Leu
      80      85      90      95
agg cag gag cac gag gag cgc ctg cag aag ata agc ctg gag cac ctg 560
Arg Gln Glu His Glu Glu Arg Leu Gln Lys Ile Ser Leu Glu His Leu
10 gag cag ctg gcc cag tac ctc ctg gca gag gag cct cac gtg gag cca 608
Glu Gln Leu Ala Gln Tyr Leu Leu Ala Glu Glu Pro His Val Glu Pro
      115      120      125
gct gga gag agg gag ctt gag gcg aag gca cgg cct cag agc tcc tgt 656
15 Ala Gly Glu Arg Glu Leu Glu Ala Lys Ala Arg Pro Gln Ser Ser Cys
      130      135      140
gac atg gag gag aag gag gag gca gca gct gat cag taaacggggcc 702
Asp Met Glu Glu Lys Glu Glu Ala Ala Ala Asp Gln
      145      150      155
20 atcctacccg aaaaaaaaaa aaaaa 727

<210> 24
<211> 470
<212> DNA
25 <213> Homo sapiens

<220>
<221> CDS
<222> 137..454
30
<220>
<221> sig_peptide
<222> 137..187
<223> Von Heijne matrix
35 score 10.7019149919754
seq VLMLLAVLIWTGA/EN

<400> 24
atcctgtgaa ctacccaaaa ggaggaaaac gaacgcagct gagcatggga tgccatataa 60
40 aaatcactta aaccagtcgc cactccttgt ttcttgagtt gtcctgtgct ggaggtctgc 120
tcagacgaag gtctcc atg gcg tta gaa gtc ttg atg ctc ctc gct gtc ttg 172
Met Ala Leu Glu Val Leu Met Leu Leu Ala Val Leu
      -15      -10
att tgg acc ggt gct gag aac ctc cat gtg aaa ata agt tgc tct ctg 220
45 ile Trp Thr Gly Ala Glu Asn Leu His Val Lys Ile Ser Cys Ser Leu
      -5      1      5      10
gac tgg ttg atg gtc tca gtt atc cca gtt gca gaa agc aga aat ctg 268
Asp Trp Leu Met Val Ser Val Ile Pro Val Ala Glu Ser Arg Asn Leu
      15      20      25
50 tat ata ttt gcg gat gaa tta cat ctg gga atg ggc tgc cct gca aat 316
Tyr Ile Phe Ala Asp Glu Leu His Leu Gly Met Gly Cys Pro Ala Asn
      30      35      40
cgg ata cat aca tat gta tat gag ttt ata tat ctt gtt cgt gat tgt 364
Arg Ile His Thr Tyr Val Tyr Glu Phe Ile Tyr Leu Val Arg Asp Cys
55
      45      50      55
ggc atc agg aca agg gta aga aca gtg att gtc tgt aaa aaa tac tgc 412
Gly Ile Arg Thr Arg Val Arg Thr Val Ile Val Cys Lys Lys Tyr Cys
      60      65      70      75
atg ttt tgt cag act ttt atg cct agt att aaa att gtc ttt 454
60 Met Phe Cys Gln Thr Phe Met Pro Ser Ile Lys Ile Val Phe
      80      85
taaaaaaaaa aaaaaa 470

```

<210> 25
 <211> 987
 <212> DNA
 <213> Homo sapiens
 5
 <220>
 <221> CDS
 <222> 238..609
 10 <220>
 <221> sig_peptide
 <222> 238..291
 <223> Von Heijne matrix
 score 10.0374888212272
 15 seq LLLLVMALPPGTT/GV
 <400> 25
 attccattca cagactcttg ttgggcagca gccacccgct cacctccatc cccaggactt 60
 agagggacgc agggcggttg gaacagagga cactccaggc gctgaccctg ggaggccagg 120
 20 accagggcca aagtcccgtg ggcaagagga gtcctcagag gtccttcatt cagcggttcc 180
 gggagggtctg ggaagcccac ggcctggctg gggcagggtc aacgccgcca ggccgcc 237
 atg gtc ctg tgc tgg ctg ctg ctt ctg gtg atg gct ctg ccc cca ggc 285
 Met Val Leu Cys Trp Leu Leu Leu Leu Val Met Ala Leu Pro Pro Gly
 -15 -10 -5
 25 acg acg ggc gtc aag gac tgc gtc ttc tgt gag ctc acc gac tcc atg 333
 Thr Thr Gly Val Lys Asp Cys Val Phe Cys Glu Leu Thr Asp Ser Met
 1 5 10
 cag tgt cct ggt acc tac atg cac tgt ggc gat gac gag gac tgc ttc 381
 Gln Cys Pro Gly Thr Tyr Met His Cys Gly Asp Asp Glu Asp Cys Phe
 15 20 25 30
 30 aca ggc cac ggg gtc gcc ccg ggc act ggt ccg gtc atc aac aaa ggc 429
 Thr Gly His Gly Val Ala Pro Gly Thr Gly Pro Val Ile Asn Lys Gly
 35 40 45
 tgc ctg cga gcc acc agc tgc ggc ctt gag gaa ccc gtc agc tac agg 477
 35 Cys Leu Arg Ala Thr Ser Cys Gly Leu Glu Glu Pro Val Ser Tyr Arg
 50 55 60
 ggc gtc acc tac agc ctc acc acc aac tgc tgc acc ggc cgc ctg tgt 525
 Gly Val Thr Tyr Ser Leu Thr Thr Asn Cys Cys Thr Gly Arg Leu Cys
 65 70 75
 40 aac aga gcc ccg agc agc cag aca gtg ggg gcc acc acc agc ctg gca 573
 Asn Arg Ala Pro Ser Ser Gln Thr Val Gly Ala Thr Thr Ser Leu Ala
 80 85 90
 ctg ggg ctg ggt atg ctg ctt cct cca cgt ttg ctg tgaccaacag 619
 Leu Gly Leu Gly Met Leu Leu Pro Pro Arg Leu Leu
 95 100 105
 45 ggaggacagg gcctgggact gttctcccag atccgccact ccccatgtcc ccatgtcctt 679
 ccccccactaa atggccagag aggccctgga caacctcttg cggccctggc ttcattccctt 739
 ctaagggtgt ccaccaggag cccggtgcta ggggaagcat cccaggcct gactgagcgg 799
 caggggagca cggcccgtgg gtttgattgt attactctgt tccactggtt ctaagacgca 859
 50 gagcttctca catctcaatc aggatgcttc tctccattgg tagcacttta gattccatga 919
 aatatggtaa aaaatatata tatatcataa taaatgacag ctgatgttca tggaaaaaaaa 979
 aaaaaaaaa 987
 <210> 26
 55 <211> 908
 <212> DNA
 <213> Homo sapiens
 <220>
 60 <221> CDS
 <222> 80..862
 <220>

```

<221> sig_peptide
<222> 80..127
<223> Von Heijne matrix
      score 3.66725851505537
5      seq FSLLSISGPPISS/SA

<400> 26
gaatgtttat cctctggaca aaccagccag cctctccaga gcaggcgtgt gatctctgta      60
ccccgcagtg ggtcagaat atg gag aac ttc tca ctc ctc agc atc tct gga      112
10      Met Glu Asn Phe Ser Leu Leu Ser Ile Ser Gly
           -15                      -10
      cct cca atc tct tcc tcc gcc ctg agt gct ttt ccc gac att atg ttc      160
      Pro Pro Ile Ser Ser Ser Ala Leu Ser Ala Phe Pro Asp Ile Met Phe
           -5                      1                      5                      10
15      tct cgt gcc acc agc ctg cca gac att gca aag aca gca gta ccc act      208
      Ser Arg Ala Thr Ser Leu Pro Asp Ile Ala Lys Thr Ala Val Pro Thr
           15                      20                      25
      gag gca tcc agc cca gct cag gcc ctg cca ccc cag tac caa agc atc      256
      Glu Ala Ser Ser Pro Ala Gln Ala Leu Pro Pro Gln Tyr Gln Ser Ile
20      att gtc agg caa ggg ata cag aac aca gtg ctc tca cca gac tgc agc      304
      Ile Val Arg Gln Gly Ile Gln Asn Thr Val Leu Ser Pro Asp Cys Ser
           30                      35                      40
      ttg ggg gac acc cag cac gga gag aag ctg agg cgg aac tgc act atc      352
      Leu Gly Asp Thr Gln His Gly Glu Lys Leu Arg Arg Asn Cys Thr Ile
25      tac cgg ccc tgg ttc tcc ccc tac agc tac ttc gtg tgt gca gac aaa      400
      Tyr Arg Pro Trp Phe Ser Pro Tyr Ser Tyr Phe Val Cys Ala Asp Lys
           45                      50                      55
30      gag agc cag ctg gag gcc tat gac ttc cca gag gtg cag cag gat gag      448
      Glu Ser Gln Leu Glu Ala Tyr Asp Phe Pro Glu Val Gln Gln Asp Glu
           60                      65                      70                      75
      ggc aag tgg gac aac tgc ctt tct gag gac atg gct gag aac atc tgt      496
      Gly Lys Trp Asp Asn Cys Leu Ser Glu Asp Met Ala Glu Asn Ile Cys
35      tcg tcc tct tcc tcc cca gag aac act tgc cct cga gaa gcc acc aag      544
      Ser Ser Ser Ser Ser Pro Glu Asn Thr Cys Pro Arg Glu Ala Thr Lys
           80                      85                      90
40      aaa tcc agg cat ggc ctg gac tcc atc aca tcc cag gac atc cta atg      592
      Lys Ser Arg His Gly Leu Asp Ser Ile Thr Ser Gln Asp Ile Leu Met
           95                      100                      105
      gct tcc aga tgg cac cca gca cag cag aat ggc tac aag tgc gtg gcc      640
      Ala Ser Arg Trp His Pro Ala Gln Gln Asn Gly Tyr Lys Cys Val Ala
           110                      115                      120
45      tgc tgc cgc atg tac ccc acc ctg gac ttc ctc aag agc cac atc aag      688
      Cys Cys Arg Met Tyr Pro Thr Leu Asp Phe Leu Lys Ser His Ile Lys
           125                      130                      135
      agg ggc ttc agg gag ggc ttc agc tgc aag gtg tac tac cgc aag ctc      736
      Arg Gly Phe Arg Glu Gly Phe Ser Cys Lys Val Tyr Tyr Arg Lys Leu
50      aaa gcc ctc tgg agc aag gag cag aag gcc cgg ctg gga gac agg ctc      784
      Lys Ala Leu Trp Ser Lys Glu Gln Lys Ala Arg Leu Gly Asp Arg Leu
           140                      145                      150                      155
      tcc tcc ggc agc tgc cag gcc ttc aat agt cct gct gaa cac ctt agg      832
      Ser Ser Gly Ser Cys Gln Ala Phe Asn Ser Pro Ala Glu His Leu Arg
55      caa att ggc ggt gaa gcc tac tta tgt ctc tagagagatg ccaataaagt      882
      Gln Ile Gly Gly Glu Ala Tyr Leu Cys Leu
           160                      165                      170                      175
60      tagtcacagc caaaaaaaaaa aaaaaa      908
           180                      185                      190                      195
           200                      205                      210                      215
           220                      225                      230                      235
           240                      245

```

```

<210> 27
<211> 762

```

<212> DNA
<213> Homo sapiens

<220>

5 <221> CDS
<222> 83..310

<220>

<221> sig_peptide
10 <222> 83..157
<223> Von Heijne matrix
score 4.72955689475746
seq LCALLSNFCPSTT/VK

15 <400> 27
ttttttctac tacaaacgcc atgggggatgc gggctctggga acagcggaaa accctaccct 60
gccctgaaaa gtccttggt ca atg tgc atg tcc ctt tct atg aaa gtt cct 112
Met Cys Met Ser Leu Ser Met Lys Val Pro
-25 -20
20 tgc tgc cta tgc gcc ttg ctc tct aac ttc tgt ccc tcc aca act gtg 160
Cys Cys Leu Cys Ala Leu Ser Asn Phe Cys Pro Ser Thr Thr Val
-15 -10 -5 1
aaa gga gac gtc gtg act tcc ttc ttt cgt gct gac tat gac tta gcc 208
Lys Gly Asp Val Val Thr Ser Phe Phe Arg Ala Asp Tyr Asp Leu Ala
25 5 10 15
agt agg tct gca gat cag tcc tcc cag aaa gtg aag ttg cgc atg ttc 256
Ser Arg Ser Ala Asp Gln Ser Ser Gln Lys Val Lys Leu Arg Met Phe
20 25 30
act ggg cgt ctt ccc atc ggc ccc ttc gcc agt gtg ggg aac gcg gcg 304
30 Thr Gly Arg Leu Pro Ile Gly Pro Phe Ala Ser Val Gly Asn Ala Ala
35 40 45
gag ctg tgagccggcg actcgggtcc ctgaggtctg gattctttct ccgctactga 360
Glu Leu
50
35 gacacggcgg acacacacaa acacagaacc acacagccag tcccaggagc ccagtaatgg 420
agagccccc aaagaagaac cagcagctga aagtcgggat cctacacctg ggcagcagac 480
agaagaagat caggatacag ctgagatccc agtcgcgcgac atggaagggtg atctgcaaga 540
gctgcatcag tcaaacaccg gggataaatc tggatttggg ttccggcgct aaggtgaaga 600
taataacctaa agaggaacac tgtaaaatgc cagaagcagg tgaagagcaa ccacaagttt 660
40 aaatgaagac aagctgaaac aacgcaagct ggttttatat tagatatttg acttaaacta 720
tctcaataaa gttttgcagc tttcaccaaa aaaaaaaaaa aa 762

<210> 28

<211> 1102

45 <212> DNA
<213> Homo sapiens

<220>

<221> CDS
50 <222> 310..906

<220>

<221> sig_peptide
<222> 310..357
55 <223> Von Heijne matrix
score 11.0931109030915
seq FPLLLSLGLVLA/EA

<400> 28

60 atacagtgc ctagagcagg catgggtggg tcacaggctt tggagagcac tctctgtcct 60
gatcttttca gttgagagac ttcagctgtt cattgtctcat ttggacttag ttcaagggtca 120
tgtcaaagaa gaagggtgcac ttacgctagt tgtagctct gtcttttgta accatcaagt 180
tccatgcgat tgatcagatt taggaggggg cgtaggggga taatcaattt tgggtgtcac 240


```

caggtaaaca gagccctcag catctgaata gaaactgaac aggaacagaa gagattcact 300
acatctgag atg gag acc ttt cct ctg ctg ctg ctc agc ctg ggc ctg gtt 351
      Met Glu Thr Phe Pro Leu Leu Leu Leu Ser Leu Gly Leu Val
      -15          -10          -5
5  ctt gca gaa gca tca gaa agc aca atg aag ata att aaa gaa gaa ttt 399
   Leu Ala Glu Ala Ser Glu Ser Thr Met Lys Ile Ile Lys Glu Glu Phe
      1          5          10
   aca gac gaa gag atg caa tat gac atg gca aaa agt ggc caa gaa aaa 447
   Thr Asp Glu Glu Met Gln Tyr Asp Met Ala Lys Ser Gly Gln Glu Lys
10 15          20          25          30
   cag acc att gag ata tta atg aac ccg atc ctg tta gtt aaa aat acc 495
   Gln Thr Ile Glu Ile Leu Met Asn Pro Ile Leu Leu Val Lys Asn Thr
      35          40          45
15 agc ctc agc atg tcc aag gat gat atg tct tcc aca tta ctg aca ttc 543
   Ser Leu Ser Met Ser Lys Asp Asp Met Ser Ser Thr Leu Leu Thr Phe
      50          55          60
   aga agt tta cat tat aat gac ccc aag gga aac agt tgc ggt aat gac 591
   Arg Ser Leu His Tyr Asn Asp Pro Lys Gly Asn Ser Ser Gly Asn Asp
      65          70          75
20 aaa gag tgt tgc aat gac atg aca gtc tgg aga aaa gtt tca gaa gca 639
   Lys Glu Cys Cys Asn Asp Met Thr Val Trp Arg Lys Val Ser Glu Ala
      80          85          90
   aac gga tgc tgc aag tgg agc aat aac ttc atc cgc agc tcc aca gaa 687
   Asn Gly Ser Cys Lys Trp Ser Asn Asn Phe Ile Arg Ser Ser Thr Glu
25 95          100          105          110
   gtg atg cgc agg gtc cac agg gcc ccc agc tgc aag ttt gta cag aat 735
   Val Met Arg Arg Val His Arg Ala Pro Ser Cys Lys Phe Val Gln Asn
      115          120          125
30 cct ggc ata agc tgc tgt gag agc cta gaa ctg gaa aat aca gtg tgc 783
   Pro Gly Ile Ser Cys Cys Glu Ser Leu Glu Leu Glu Asn Thr Val Cys
      130          135          140
   cag ttc act aca ggc aaa caa ttc ccc agg tgc caa tac cat agt gtt 831
   Gln Phe Thr Thr Gly Lys Gln Phe Pro Arg Cys Gln Tyr His Ser Val
      145          150          155
35 acc tca tta gag aag ata ttg aca gtg ctg aca ggt cat tct ctg atg 879
   Thr Ser Leu Glu Lys Ile Leu Thr Val Leu Thr Gly His Ser Leu Met
      160          165          170
   agc tgg tta gtt tgt ggc tct aag ttg taaatccac agagctttag 926
   Ser Trp Leu Val Cys Gly Ser Lys Leu
40 175          180
   gactagggtc ttactaaaga aggacctctt cttgttcatt cttgtttaaa cctttcctta 986
   atatctactc ttttagcacta tagtgaactc ctgattatatt attctaactg gaggagtga 1046
   aaatccaaaaa ttgtggataa ttcaattaaa agttatgact gaaaaaaaaa aaaaaa 1102

45 <210> 29
   <211> 436
   <212> DNA
   <213> Homo sapiens

50 <220>
   <221> CDS
   <222> 24..287

   <220>
55 <221> sig_peptide
   <222> 24..131
   <223> Von Heijne matrix
       score 3.79790641648006
       seq ILMRDFSPSGIFG/AF

60 <400> 29
   acagcggaca ccaggactcc aaa atg gcg tca gtt gta cca gtg aag gac aag 53
       Met Ala Ser Val Val Pro Val Lys Asp Lys

```

33

	cag ggc tcc ttt gat gtc agt gtc aag ggc atc agc att tcg gtc aac	554
	Gln Gly Ser Phe Asp Val Ser Val Lys Gly Ile Ser Ile Ser Val Asn	
	105 110 115	
	ctc ctg ttg ggc agc gat tcc tcc ggg agg ccc aca gtt act gcc tcc	602
5	Leu Leu Leu Gly Ser Asp Ser Ser Gly Arg Pro Thr Val Thr Ala Ser	
	120 125 130	
	agc tgc agc agt gac atc gct gac gtg gag gtg gac atg tcg gga gac	650
	Ser Cys Ser Ser Asp Ile Ala Asp Val Glu Val Asp Met Ser Gly Asp	
	135 140 145	
10	ttg ggg tgg ctg ttg aac ctc ttc cac aac cag att gag tcc aag ttc	698
	Leu Gly Trp Leu Leu Asn Leu Phe His Asn Gln Ile Glu Ser Lys Phe	
	150 155 160	
	cag aaa gta ctg gag agc agg att tgc gaa atg atc cag aaa tcg gtg	746
	Gln Lys Val Leu Glu Ser Arg Ile Cys Glu Met Ile Gln Lys Ser Val	
15	165 170 175 180	
	tcc tcc gat cta cag cct tat ctc caa act ctg aca gtt aca aca gag	794
	Ser Ser Asp Leu Gln Pro Tyr Leu Gln Thr Leu Thr Val Thr Thr Glu	
	185 190 195	
20	att gac agt ttc gcc gac att gat tat agc tta gtg gaa gcc cct cgg	842
	Ile Asp Ser Phe Ala Asp Ile Asp Tyr Ser Leu Val Glu Ala Pro Arg	
	200 205 210	
	gca aca gcc cag atg ctg gag gtg atg ttt aag ggt gaa atc ttt cat	890
	Ala Thr Ala Gln Met Leu Glu Val Met Phe Lys Gly Glu Ile Phe His	
	215 220 225	
25	cgt aac cac cgt tct cca gtt acc ctc ctt gct gca gtc atg agc ctt	938
	Arg Asn His Arg Ser Pro Val Thr Leu Leu Ala Ala Val Met Ser Leu	
	230 235 240	
	cct gag gaa cac aac aaa atg gtc tac ttt gcc atc tcg gat tat gtc	986
	Pro Glu Glu His Asn Lys Met Val Tyr Phe Ala Ile Ser Asp Tyr Val	
30	245 250 255 260	
	ttc aac acg gcc agc ctg gtt tat cat gag gaa gga tat ctg aac ttc	1034
	Phe Asn Thr Ala Ser Leu Val Tyr His Glu Glu Gly Tyr Leu Asn Phe	
	265 270 275	
35	tcc atc aca gat gac atg ata ccg cct gac tct aat atc cga ctg acc	1082
	Ser Ile Thr Asp Asp Met Ile Pro Pro Asp Ser Asn Ile Arg Leu Thr	
	280 285 290	
	acc aag tcc ttc cga ccc ttc gtc cca cgg tta gcc agg ctc tac ccc	1130
	Thr Lys Ser Phe Arg Pro Phe Val Pro Arg Leu Ala Arg Leu Tyr Pro	
	295 300 305	
40	aac atg aac ctg gaa ctc cag gga tca gtg ccc tct gct ccg ctc ctg	1178
	Asn Met Asn Leu Glu Leu Gln Gly Ser Val Pro Ser Ala Pro Leu Leu	
	310 315 320	
	aac ttc agc cct ggg aat ctg tct gtg gac ccc tat atg gag ata gat	1226
	Asn Phe Ser Pro Gly Asn Leu Ser Val Asp Pro Tyr Met Glu Ile Asp	
45	325 330 335 340	
	gcc ttt gtg ctc ctg ccc agc tcc agc aag gag cct gtc ttc cgg ctc	1274
	Ala Phe Val Leu Leu Pro Ser Ser Ser Lys Glu Pro Val Phe Arg Leu	
	345 350 355	
50	agt gtg gcc act aat gtg tcc gcc acc ttg acc ttc aat acc agc aag	1322
	Ser Val Ala Thr Asn Val Ser Ala Thr Leu Thr Phe Asn Thr Ser Lys	
	360 365 370	
	atc act ggg ttc ctg aag cca gga aag gta aaa gtg gaa ctg aaa gaa	1370
	Ile Thr Gly Phe Leu Lys Pro Gly Lys Val Lys Val Glu Leu Lys Glu	
	375 380 385	
55	tcc aaa gtt gga cta ttc aat gca gag ctg ttg gaa gcg ctc ctc aac	1418
	Ser Lys Val Gly Leu Phe Asn Ala Glu Leu Leu Glu Ala Leu Leu Asn	
	390 395 400	
	tat tac atc ctt aac acc ttc tac ccc aag ttc aat gat aag ttg gcc	1466
	Tyr Tyr Ile Leu Asn Thr Phe Tyr Pro Lys Phe Asn Asp Lys Leu Ala	
60	405 410 415 420	
	gaa ggc ttc ccc ctt cct ctg ctg aag cgt gtt cag ctc tac gac ctt	1514
	Glu Gly Phe Pro Leu Pro Leu Leu Lys Arg Val Gln Leu Tyr Asp Leu	
	425 430 435	

```

ggg ctg cag atc cat aag gac ttc ctg ttc ttg ggt gcc aat gtc caa      1562
Gly Leu Gln Ile His Lys Asp Phe Leu Phe Leu Gly Ala Asn Val Gln
      440                      445                      450
tac atg aga gtt tgaggacaag aaagatgaag cttggagggtc acagctggat      1614
5 Tyr Met Arg Val
      455
ctgcttggttg catttccagc tgtgcagcac gtctcagaga ttcttgaaga atgaagacat      1674
ttctgctctc agctccgggg gtgaggtgtg cctggcctct gcctccaccc tcctcctctt      1734
caccaggtgc atgcatgccc tctctgagtc tggactttgc ttcccccca ggagggacca      1794
10 ccctccctga ctggcctggg atatctttac aagcaggcac tgtatttttt tattcgccat      1854
ctgatcccca tgcctagcag agtgctggca cttagtaggt cctcaataaa tatttattaa      1914
atgatgacaa aaaaaaaaaa aaaa      1938

<210> 31
15 <211> 1116
    <212> DNA
    <213> Homo sapiens

    <220>
20 <221> CDS
    <222> 117..545

    <220>
    <221> sig_peptide
25 <222> 117..245
    <223> Von Heijne matrix
        score 5.65876793443964
        seq VVSFALIATLVYA/LF

30 <400> 31
ataagggggac gtctagtggg ttgcccggga ggggtggcgg gagcggtcct ggaaataatc      60
tgtcctctgt cgccgggaac tggcgaggta gttccttcgc ggtggagaga cctgga atg      119
                                     Met
gcc aaa tat caa ggt gaa gtt caa agt ttg aaa ctg gat gat gat tca      167
35 Ala Lys Tyr Gln Gly Glu Val Gln Ser Leu Lys Leu Asp Asp Asp Ser
    -40                      -35                      -30
gtt ata gaa gga gta agc gac caa gta ctt gtg gca gtt gtg gtc agt      215
Val Ile Glu Gly Val Ser Asp Gln Val Leu Val Ala Val Val Val Ser
    -25                      -20                      -15
40 ttc gct ttg att gct acc ctg gta tat gca ctt ttc aga aat gta cat      263
Phe Ala Leu Ile Ala Thr Leu Val Tyr Ala Leu Phe Arg Asn Val His
    -10                      -5                      1                      5
caa aac att cac cca gaa aac cag gag cta gta agg gta ctt cga gaa      311
Gln Asn Ile His Pro Glu Asn Gln Glu Leu Val Arg Val Leu Arg Glu
45                                10                      15                      20
cag ctt caa aca gaa cag gat gca cct gct gcc act cga cag cag ttc      359
Gln Leu Gln Thr Glu Gln Asp Ala Pro Ala Ala Thr Arg Gln Gln Phe
    25                      30                      35
tac act gac atg tac tgt ccc atc tgc ctg cac caa gcc tcc ttc ccg      407
50 Tyr Thr Asp Met Tyr Cys Pro Ile Cys Leu His Gln Ala Ser Phe Pro
    40                      45                      50
gtg gag acc aac tgt gga cat ctt ttt tgt ggt gcc tgc att att gct      455
Val Glu Thr Asn Cys Gly His Leu Phe Cys Gly Ala Cys Ile Ile Ala
    55                      60                      65                      70
55 tac tgg cga tat ggt tca tgg ctt ggg gca atc agt tgt cca atc tgt      503
Tyr Trp Arg Tyr Gly Ser Trp Leu Gly Ala Ile Ser Cys Pro Ile Cys
    75                      80                      85
aga caa acg aga cat ggc cac att gca ttg tcc aga aca gct      545
Arg Gln Thr Arg His Gly His Ile Ala Leu Ser Arg Thr Ala
60                                90                      95                      100
tagaccatga cagttagcat cgaagccacc tgaggagggg ggcagtaacc ttactcctaa      605
cagtatttgg tgaagatgat cagtctcagg atgttctgag attgcatcag gatattaatg      665
attataaccg gagattctca gggcaaccca gatctgtaag taatgctaaa gcatgttcaa      725

```

	agtttagagga	agacacattt	cttctctttt	gtaaagttag	gtttaccaac	aagtattctt	785
	tgactatgag	aaatcttggc	caggcacagt	agctaacgcc	tataatccta	gcactttggg	845
	aggccaaggc	aggtggatca	cttgagccca	ggagtttgag	accagccttg	gaaacatgat	905
	gaaaccccat	ctctagaaaa	aacacccaaa	aattggacaa	gagtgttggc	acatgcctgt	965
5	agtccctgct	tcttgggagg	ctgaaatggg	aggatcacct	gagcccagga	ggttgaggct	1025
	atagttagcc	atgatcgac	tattgcactc	ccacctgggt	ggcagtgaga	cccttcctca	1085
	aaaaacaaga	aaagaaaaaa	aaaaaaaaaa	a			1116
	<210> 32						
10	<211> 1114						
	<212> DNA						
	<213> Homo sapiens						
	<220>						
15	<221> CDS						
	<222> 117..362						
	<400> 32						
20	ataagggggac	gtctagtggg	ttgcccggga	ggggtggcgg	gagcggtcct	ggaaataatc	60
	tgctctctgt	cgccgggaac	tggcgaggta	gttccttcgc	ggtggagaga	cctgga atg	119
						Met	
						1	
	gcc aaa tat	caa ggt gaa	gtt caa agt	ttg aaa ctg	gat gat gat	tca	167
25	Ala Lys Tyr	Gln Gly Glu	Val Gln Ser	Leu Lys Leu	Asp Asp Asp	Ser	
		5	10	15			
	gtt ata gaa	gga gta agc	gac caa gta	ctt gtg gca	gtt gtg gtc	agt	215
	Val Ile Glu	Gly Val Ser	Asp Gln Val	Leu Val Ala	Val Val Val	Ser	
		20	25	30			
	ttc gct ttg	att gct acc	ctg gta tat	gca ctt ttc	aga aat gta	cat	263
30	Phe Ala Leu	Ile Ala Thr	Leu Val Tyr	Ala Leu Phe	Arg Asn Val	His	
		35	40	45			
	caa aac att	cac cca gaa	aac cag gag	cta gta agg	gta ctt cga	gaa	311
	Gln Asn Ile	His Pro Glu	Asn Gln Glu	Leu Val Arg	Val Leu Arg	Glu	
		50	55	60	65		
35	cag ctt caa	aca gaa cag	gat gca cct	gct gac tcg	aca gca gtt	cta	359
	Gln Leu Gln	Thr Glu Gln	Asp Ala Pro	Ala Asp Ser	Thr Ala Val	Leu	
		70	75	80			
	cac tgacatgtac	tgtcccatct	gcctgcacca	agcctccttc	ccggtggaga		412
	His						
40	ccaactgtgg	acatcttttt	tgtgggtgct	gcattattgc	ttactggcga	tatggttcat	472
	ggcttggggc	aatcagttgt	ccaatctgta	gacaaacgag	acatggccac	attgcattgt	532
	ccagaacagc	ttagaccatg	acagtttagca	tcgaagccac	ctgaggaggg	aggcagtaac	592
	cttactccta	acagtatttg	gtgaagatga	tcagtctcag	gatgttctga	gattgcatca	652
	ggatattaat	gattataacc	ggagattctc	agggcaaccc	agatctgtaa	gtaatgctaa	712
45	agcatgttca	aagtttagagg	aagacacatt	tcttctcttt	tgtaaagtga	ggtttaccac	772
	caagtattct	ttgactatga	gaaatcttgg	ccaggcacag	tagctaacgc	ctataatcct	832
	agcacttttg	gaggccaagg	cagggtggatc	acttgagccc	aggagtttga	gaccagcctt	892
	ggaaacatga	tgaaacccca	tctctagaaa	aaacacccaa	aaattggaca	agagtgttgg	952
	cacatgcctg	tagtccctgc	ttcttggggag	gctgaaatgg	gaggatcacc	tgagcccagg	1012
50	aggttgaggc	tatagttagc	catgatcgca	ctattgcact	cccacctggg	tggcagttag	1072
	acccttcctc	aaaaaacaag	aaaagaaaaa	aaaaaaaaaa	aa		1114
	<210> 33						
	<211> 2072						
55	<212> DNA						
	<213> Homo sapiens						
	<220>						
	<221> CDS						
60	<222> 144..1262						
	<220>						
	<221> sig_peptide						

<222> 144..224

<223> Von Heijne matrix

score 5.14258625256317

seq FLCQRLVLSTLSG/RP

5

<400> 33

	acgtggacgc	gtctgggctg	ctggaggcag	cccagaccgc	cgccgctcgg	gtcgccgcc	60
	ccaccacat	cggagtcacg	agtcccgct	ctgtccgaag	tcgccgctct	cgggctgctc	120
10	acgtctcttc	ggagagcgcg	cac atg	gcg act	cag gcg	cac tcc	173
			Met Ala Thr Gln Ala His Ser Leu Ser Tyr				
			-25			-20	
	gca ggg tgc aac ttc ttg tgc caa cgt ctg gtc ctg tct acc ctg agc	221					
	Ala Gly Cys Asn Phe Leu Cys Gln Arg Leu Val Leu Ser Thr Leu Ser						
	-15	-10	-5				
15	ggg cgc ccc gtc aaa atc cga aag att cgg gcc aga gac gac aac ccg	269					
	Gly Arg Pro Val Lys Ile Arg Lys Ile Arg Ala Arg Asp Asp Asn Pro						
	1	5	10	15			
	ggc ctc cga gat ttt gaa gcc agc ttc ata agg cta ttg gac aaa ata	317					
	Gly Leu Arg Asp Phe Glu Ala Ser Phe Ile Arg Leu Leu Asp Lys Ile						
20		20	25	30			
	acg aat ggt tct cga att gaa ata aac caa aca gga aca acc tta tat	365					
	Thr Asn Gly Ser Arg Ile Glu Ile Asn Gln Thr Gly Thr Thr Leu Tyr						
	35	40	45				
25	tat cag cct ggc ctc ctg tat ggt gga tct gtg gaa cat gac tgt agc	413					
	Tyr Gln Pro Gly Leu Leu Tyr Gly Gly Ser Val Glu His Asp Cys Ser						
	50	55	60				
	gtc ctt cgt ggc att ggg tat tac ctg gag agt ctt ctt tgc ttg gct	461					
	Val Leu Arg Gly Ile Gly Tyr Tyr Leu Glu Ser Leu Leu Cys Leu Ala						
	65	70	75				
30	cca ttt atg aag cac ccg tta aaa ata gtt cta cga gga gtg acc aat	509					
	Pro Phe Met Lys His Pro Leu Lys Ile Val Leu Arg Gly Val Thr Asn						
	80	85	90	95			
	gat cag att gac cct tca gtt gat gtt ctt aag gca aca gca ctc cct	557					
	Asp Gln Ile Asp Pro Ser Val Asp Val Leu Lys Ala Thr Ala Leu Pro						
35		100	105	110			
	ttg ttg aaa caa ttt ggg att gat ggt gaa tca ttt gaa ctg aag att	605					
	Leu Leu Lys Gln Phe Gly Ile Asp Gly Glu Ser Phe Glu Leu Lys Ile						
	115	120	125				
	gtg cga cgg gga atg cct ccc gga gga gga ggc gaa gtg gtt ttc tca	653					
40	Val Arg Arg Gly Met Pro Pro Gly Gly Gly Glu Val Val Phe Ser						
	130	135	140				
	tgt cct gtg agg aag gtc ttg aag ccc att caa ctc aca gat cca gga	701					
	Cys Pro Val Arg Lys Val Leu Lys Pro Ile Gln Leu Thr Asp Pro Gly						
	145	150	155				
45	aaa atc aaa cgt att aga gga atg gcg tac tct gta cgt gtg tca cct	749					
	Lys Ile Lys Arg Ile Arg Gly Met Ala Tyr Ser Val Arg Val Ser Pro						
	160	165	170	175			
	cag atg gcg aac cgg att gtg gat tct gca agg agc atc ctc aac aag	797					
	Gln Met Ala Asn Arg Ile Val Asp Ser Ala Arg Ser Ile Leu Asn Lys						
50		180	185	190			
	ttc ata cct gat atc tat att tac aca gat cac att aaa gga gtc aac	845					
	Phe Ile Pro Asp Ile Tyr Ile Tyr Thr Asp His Ile Lys Gly Val Asn						
	195	200	205				
	tct ggg aag tct ccg ggc ttt ggg ttg tca ctg gtt gct gag acc acc	893					
55	Ser Gly Lys Ser Pro Gly Phe Gly Leu Ser Leu Val Ala Glu Thr Thr						
	210	215	220				
	agt ggc acc ttc ctc agt gct gaa ctg gcc tcc aac ccc cag ggc cag	941					
	Ser Gly Thr Phe Leu Ser Ala Glu Leu Ala Ser Asn Pro Gln Gly Gln						
	225	230	235				
60	gga gca gca gta ctt cca gag gac ctt ggc agg aac tgt gcc cgg ctg	989					
	Gly Ala Ala Val Leu Pro Glu Asp Leu Gly Arg Asn Cys Ala Arg Leu						
	240	245	250	255			
	ctg ctg gag gaa atc tac agg ggt gga tgc gta gac tcg acc aac caa	1037					

```

Leu Leu Glu Glu Ile Tyr Arg Gly Gly Cys Val Asp Ser Thr Asn Gln
      260      265      270
agc ctg gcg cta cta ctc atg acc ctt gga cag cag gat gtt tcc aaa 1085
Ser Leu Ala Leu Leu Leu Met Thr Leu Gly Gln Gln Asp Val Ser Lys
5      275      280      285
gtc ctg cta ggc cct ctc tct ccc tac acg ata gaa ttt ttg cgg cat 1133
Val Leu Leu Gly Pro Leu Ser Pro Tyr Thr Ile Glu Phe Leu Arg His
      290      295      300
ttg aag agc ttt ttc cag att atg ttt aaa att gaa acc aag cca tgt 1181
10 Leu Lys Ser Phe Phe Gln Ile Met Phe Lys Ile Glu Thr Lys Pro Cys
      305      310      315
ggg gaa gaa ctc aag ggt ggg gat aaa gtg ctg atg acc tgt gtt ggc 1229
Gly Glu Glu Leu Lys Gly Gly Asp Lys Val Leu Met Thr Cys Val Gly
      320      325      330      335
15 att ggt ttc tcc aac ctt agc agg acc ctc aag tgataacccat cacaagataa 1282
Ile Gly Phe Ser Asn Leu Ser Arg Thr Leu Lys
      340      345
ggccccagtg cctacagaca aagcagaagc tgccacggac accaatggga ccaagtccaa 1342
atggattaat ccaggacaga atagccactt gcttaatttt ctgtgaagaa atatcaatat 1402
20 acaaataaaa gacatccctg tagcatatgg tttccagctg tttctccagt ggcattgcca 1462
ttgcccagga ggggccagtg caccatgaga gctcccttgc cttacctgga ggaagaatgt 1522
gccttcaggc cacagtcgtg ctgctagaac agtctcgtag ctgcagttca gctgtgcttc 1582
ctcagcctac tatcataggg ttccctcagcc ctctgtcata tggctgtttt gcaaacctgt 1642
ggagtctgtt actgtttctt ctgcaaggac tcacctcctt gagccttggt tttgtgtgta 1702
25 gggattaaat gagataatat gagtggcagc tcttcatgag tcctgcagtg ctaagcaaatt 1762
gtcagaaatt ggtgtattag actattttatc tttgatcttc tgaatggatt gctgtcatgg 1822
acacggacac ggatcttcat ctgggttcatt gtattttatat gtgaggggatg gatggctgcg 1882
gggctccaag taagttattg ggatgttttt atattccagg tgtgctgtac gttcttattt 1942
tattttcaca atagctctgt gatgtaagt ctatctccat gagaaaattc ataaaggggtg 2002
30 ttttgttcat ttgaaatgta taatgtaaag acattaaatc tcctcattta aggaaaaaaa 2062
aaaaaaa
<210> 34
<211> 409
35 <212> DNA
    <213> Homo sapiens
<220>
<221> CDS
40 <222> 35..316
<220>
<221> sig_peptide
<222> 35..109
45 <223> Von Heijne matrix
    score 5.38058532480537
    seq AVTSLLSPTPATA/LA
<400> 34
50 tttttttcga gaccggaagt gagtgatega aagc atg gcg tcg gtg gtg ttg gcg 55
    Met Ala Ser Val Val Leu Ala
    -25 -20
ctg agg acc cgg aca gcc gtt aca tcc ttg cta agc ccc act ccg gct 103
Leu Arg Thr Arg Thr Ala Val Thr Ser Leu Leu Ser Pro Thr Pro Ala
55 -15 -10 -5
aca gct ctt gct gtc aga tac gca tcc aag aag tcg ggt ggt agc tcc 151
Thr Ala Leu Ala Val Arg Tyr Ala Ser Lys Lys Ser Gly Gly Ser Ser
1 5 10
aaa aac ctc ggt gga aag tca tca ggc aga cgc caa ggc att aag aaa 199
60 Lys Asn Leu Gly Gly Lys Ser Ser Gly Arg Arg Gln Gly Ile Lys Lys
15 20 25 30
atg gaa ggt cac tat gtt cat gct ggg aac atc att gca aca cag cgc 247
Met Glu Gly His Tyr Val His Ala Gly Asn Ile Ile Ala Thr Gln Arg

```

```

      35      40      45
cat ttc cgc tgg cac cca ggt gcc cat gtg agt tgc tcc gtt gct gcc      295
His Phe Arg Trp His Pro Gly Ala His Val Ser Cys Ser Val Ala Ala
      50      55      60
5 ccc ctt ttt cct ttt cta ggt tgacctctcc ttgccctaa gcatggtaat      346
Pro Leu Phe Pro Phe Leu Gly
      65
aacagttgca tgtattgagt gcttaccaaaa tggcaagcat tgtgcaaaaa aaaaaaaaaa      406
aaa      409

10 <210> 35
    <211> 836
    <212> DNA
    <213> Homo sapiens

15 <220>
    <221> CDS
    <222> 177..767

20 <220>
    <221> sig_peptide
    <222> 177..236
    <223> Von Heijne matrix
        score 6.51720597568932

25 seq LAVILTLLGLAIL/AI

    <400> 35
aatctgctcc acgcaatttc tcagtgatcc tctgcatctc tgcctacaag ggctccctg      60
acacccaagt tcatattgct cagaaacagt gaacttgagt ttttcatttt acctgatct      120
30 ctctctgaca aagaaatcca gatgatgcga gacctgatga agacaatata tggaaa atg      179
Met
-20

aca gtc ttg gaa ata act ttg gct gtc atc ctg act cta ctg gga ctt      227
Thr Val Leu Glu Ile Thr Leu Ala Val Ile Leu Thr Leu Leu Gly Leu
35 -15 -10 -5
gcc atc ctg gct att ttg tta aca aga tgg gca cga cgt aag caa agt      275
Ala Ile Leu Ala Ile Leu Leu Thr Arg Trp Ala Arg Arg Lys Gln Ser
1 5 10
gaa atg tat atc tcc aga tac agt tca gaa caa agt gct aga ctt ctg      323
40 Glu Met Tyr Ile Ser Arg Tyr Ser Ser Glu Gln Ser Ala Arg Leu Leu
15 20 25
gac tat gag gat ggt aga gga tcc cga cat gca tat tca aca caa agt      371
Asp Tyr Glu Asp Gly Arg Gly Ser Arg His Ala Tyr Ser Thr Gln Ser
30 35 40 45
gag aga tcc aaa aga gat tac aca cca tca acc aac tct cta gca ctg      419
45 Glu Arg Ser Lys Arg Asp Tyr Thr Pro Ser Thr Asn Ser Leu Ala Leu
50 55 60
tct cga tca agt att gct tta cct caa gga tcc atg agt agt ata aaa      467
Ser Arg Ser Ser Ile Ala Leu Pro Gln Gly Ser Met Ser Ser Ile Lys
65 70 75
tgt tta caa aca act gaa gaa cct cct tcc aga act gca gga gcc atg      515
Cys Leu Gln Thr Thr Glu Glu Pro Pro Ser Arg Thr Ala Gly Ala Met
80 85 90
atg caa ttc aca gcc cct att ccc gga gct aca gga cct atc aag ctc      563
55 Met Gln Phe Thr Ala Pro Ile Pro Gly Ala Thr Gly Pro Ile Lys Leu
95 100 105
tct caa aaa acc att gtg caa act cta gga cct att gta caa tat cct      611
Ser Gln Lys Thr Ile Val Gln Thr Leu Gly Pro Ile Val Gln Tyr Pro
110 115 120 125
60 gga tcc aat ggg agg ata aac ata agc cag ctc acc tca gag gat ctc      659
Gly Ser Asn Gly Arg Ile Asn Ile Ser Gln Leu Thr Ser Glu Asp Leu
130 135 140
act ggg gct aaa gga agg gtc aca tct ggt cca cag ttc cct aat agc      707

```


Thr Gly Ala Lys Gly Arg Val Thr Ser Gly Pro Gln Phe Pro Asn Ser
 145 150 155
 cac cat gtg cca gag aat cta cat gga tac atg aat tcc ctt tcc ctt 755
 His His Val Pro Glu Asn Leu His Gly Tyr Met Asn Ser Leu Ser Leu
 5 160 165 170
 ttc tcc cct gct tgactccctc tcccttatgt gtaaacaatt taaaaatatg 807
 Phe Ser Pro Ala
 175
 atagtgtata aatgaaaaaa aaaaaaaaaa 836
 10
 <210> 36
 <211> 1323
 <212> DNA
 <213> Homo sapiens
 15
 <220>
 <221> CDS
 <222> 208..1239
 20 <220>
 <221> sig_peptide
 <222> 208..294
 <223> Von Heijne matrix
 score 5.73027134157378
 25 seq GLVLICVCSKTHS/LK
 <400> 36
 agtctcgtat cgcgccccggg aggcgcgcgga gccagcggc tggcgccaga tccaggctcc 60
 tggaagaacc atgtccggca gctactggtc atgccaggca cacactgctg cccaagagga 120
 30 gctgctgttt gaattatctg tgaatgttgg gaagaggaat gccagagctg cgggctgaaa 180
 attaccaac caagagaaat ctgcagg atg gac ttt ctg gtc ctc ttc ttg ttc 234
 Met Asp Phe Leu Val Leu Phe Leu Phe
 -25
 tac ctg gct tcg gtg ctg atg ggt ctt gtt ctt atc tgc gtc tgc tcg 282
 35 Tyr Leu Ala Ser Val Leu Met Gly Leu Val Leu Ile Cys Val Cys Ser
 -20 -15 -10 -5
 aaa acc cat agc ttg aaa ggc ctg gcc agg gga gca cag ata ttt 330
 Lys Thr His Ser Leu Lys Gly Leu Ala Arg Gly Gly Ala Gln Ile Phe
 1 5 10
 40 tcc tgt ata att cca gaa tgt ctt cag aga gcc gtg cat gga ttg ctt 378
 Ser Cys Ile Ile Pro Glu Cys Leu Gln Arg Ala Val His Gly Leu Leu
 15 20 25
 cat tac ctt ttc cat acg aga aac cac acc ttc att gtc ctg cac ctg 426
 45 His Tyr Leu Phe His Thr Arg Asn His Thr Phe Ile Val Leu His Leu
 30 35 40
 gtc ttg caa ggg atg gtt tat act gag tac acc tgg gaa gta ttt ggc 474
 Val Leu Gln Gly Met Val Tyr Thr Glu Tyr Thr Trp Glu Val Phe Gly
 45 50 55 60
 tac tgt cag gag ctg gag ttg tcc ttg cat tac ctt ctt ctg ccc tat 522
 50 Tyr Cys Gln Glu Leu Glu Leu Ser Leu His Tyr Leu Leu Leu Pro Tyr
 65 70 75
 ctg ctg cta ggt gta aac ctg ttt ttt ttc acc ctg act tgt gga acc 570
 Leu Leu Leu Gly Val Asn Leu Phe Phe Phe Thr Leu Thr Cys Gly Thr
 80 85 90
 55 aat cct ggc att ata aca aaa gca aat gaa tta tta ttt ctt cat gtt 618
 Asn Pro Gly Ile Ile Thr Lys Ala Asn Glu Leu Leu Phe Leu His Val
 95 100 105
 tat gaa ttt gat gaa gtg atg ttt cca aag aac gtg agg tgc tct act 666
 Tyr Glu Phe Asp Glu Val Met Phe Pro Lys Asn Val Arg Cys Ser Thr
 110 115 120
 60 tgt gat tta agg aaa cca gct cga tcc aag cac tgc agt gtg tgt aac 714
 Cys Asp Leu Arg Lys Pro Ala Arg Ser Lys His Cys Ser Val Cys Asn
 125 130 135 140

	tgg tgt gtg cac cgt ttc gac cat cac tgt gtt tgg gtg aac aac tgc	762
	Trp Cys Val His Arg Phe Asp His His Cys Val Trp Val Asn Asn Cys	
	145 150 155	
5	atc ggg gcc tgg aac atc agg tac ttc ctc atc tac gtc ttg acc ttg	810
	Ile Gly Ala Trp Asn Ile Arg Tyr Phe Leu Ile Tyr Val Leu Thr Leu	
	160 165 170	
	acg gcc tcg gct gcc acc gtc gcc att gtg agc acc act ttt ctg gtc	858
	Thr Ala Ser Ala Ala Thr Val Ala Ile Val Ser Thr Thr Phe Leu Val	
	175 180 185	
10	cac ttg gtg gtg atg tca gat tta tac cag gag act tac atc gat gac	906
	His Leu Val Val Met Ser Asp Leu Tyr Gln Glu Thr Tyr Ile Asp Asp	
	190 195 200	
	ctt gga cac ctc cat gtt atg gac acg gtc ttt ctt att cag tac ctg	954
	Leu Gly His Leu His Val Met Asp Thr Val Phe Leu Ile Gln Tyr Leu	
15	205 210 215 220	
	ttc ctg act ttt cca cgg att gtc ttc atg ctg ggc ttt gtc gtg gtt	1002
	Phe Leu Thr Phe Pro Arg Ile Val Phe Met Leu Gly Phe Val Val Val	
	225 230 235	
20	ctg agc ttc ctc ctg ggt ggc tac ctg ttg ttt gtc ctg tat ctg gcg	1050
	Leu Ser Phe Leu Leu Gly Gly Tyr Leu Leu Phe Val Leu Tyr Leu Ala	
	240 245 250	
	gcc acc aac cag act act aac gag tgg tac aga ggt gac tgg gcc tgg	1098
	Ala Thr Asn Gln Thr Thr Asn Glu Trp Tyr Arg Gly Asp Trp Ala Trp	
	255 260 265	
25	tgc cag cgt tgt ccc ctt gtg gcc tgg cct ccg tca gca gag ccc caa	1146
	Cys Gln Arg Cys Pro Leu Val Ala Trp Pro Pro Ser Ala Glu Pro Gln	
	270 275 280	
	gtc cac cgg aac att cac tcc cat ggg ctt cgg agc aac ctt caa gag	1194
	Val His Arg Asn Ile His Ser His Gly Leu Arg Ser Asn Leu Gln Glu	
30	285 290 295 300	
	atc ttt cta cct gcc ttt cca tgt cat gag agg aag aaa caa gaa	1239
	Ile Phe Leu Pro Ala Phe Pro Cys His Glu Arg Lys Lys Gln Glu	
	305 310 315	
35	tgacaagtgt atgactgcct ttgagctgta gttcccggttt atttacacat gtggatcctc	1299
	gtttttccaaa aaaaaaaaaa aaaa	1323
	<210> 37	
	<211> 1945	
	<212> DNA	
40	<213> Homo sapiens	
	<220>	
	<221> CDS	
	<222> 60..1682	
45	<220>	
	<221> sig_peptide	
	<222> 60..143	
	<223> Von Heijne matrix	
50	score 3.75144398608723	
	seq SGLLLQVLFRLIT/FV	
	<400> 37	
	atcgcgacta aacggagtggt cggcggcatt tcctggtgtc tgagcctggc gcggaggct	59
55	atg ggc agc cag gag gtg ctg ggc cac gcg gcc cgg ctg tcc tcc tcc	107
	Met Gly Ser Gln Glu Val Leu Gly His Ala Ala Arg Leu Ser Ser Ser	
	-25 -20 -15	
	ggt ctc ctc ctg cag gtg ttg ttt cgg ttg atc acc ttt gtc ttg aat	155
	Gly Leu Leu Leu Gln Val Leu Phe Arg Leu Ile Thr Phe Val Leu Asn	
60	-10 -5 1	
	gca ttt att ctt cgc ttc ctg tca aag gaa atc gtt ggc gta gta aat	203
	Ala Phe Ile Leu Arg Phe Leu Ser Lys Glu Ile Val Gly Val Val Asn	
	5 10 15 20	

	gta	aga	cta	acg	ctg	ctt	tac	tca	acc	acc	ctc	ttc	ctg	gcc	aga	gag	251
	Val	Arg	Leu	Thr	Leu	Leu	Tyr	Ser	Thr	Thr	Leu	Phe	Leu	Ala	Arg	Glu	
				25						30					35		
5	gcc	ttc	cgc	aga	gca	tgt	ctc	agt	ggg	ggc	acc	cag	cga	gac	tgg	agc	299
	Ala	Phe	Arg	Arg	Ala	Cys	Leu	Ser	Gly	Gly	Thr	Gln	Arg	Asp	Trp	Ser	
				40					45					50			
	cag	acc	ctc	aac	ctg	ctg	tgg	cta	aca	gtc	ccc	ctg	ggg	gtg	ttt	tgg	347
	Gln	Thr	Leu	Asn	Leu	Leu	Trp	Leu	Thr	Val	Pro	Leu	Gly	Val	Phe	Trp	
				55					60					65			
10	tcc	tta	ttc	ctg	ggc	tgg	atc	tgg	ttg	cag	ctg	ctt	gaa	gtg	cct	gat	395
	Ser	Leu	Phe	Leu	Gly	Trp	Ile	Trp	Leu	Gln	Leu	Leu	Glu	Val	Pro	Asp	
				70				75				80					
	cct	aat	gtt	gtc	cct	cac	tat	gca	act	gga	gtg	gtg	ctg	ttt	ggg	ctc	443
	Pro	Asn	Val	Val	Pro	His	Tyr	Ala	Thr	Gly	Val	Val	Leu	Phe	Gly	Leu	
15	85					90				95						100	
	tcg	gca	gtg	gtg	gag	ctt	cta	gga	gag	ccc	ttt	tgg	gtc	ttg	gca	caa	491
	Ser	Ala	Val	Val	Glu	Leu	Leu	Gly	Glu	Pro	Phe	Trp	Val	Leu	Ala	Gln	
					105					110						115	
20	gca	cat	atg	ttt	gtg	aag	ctc	aag	gtg	att	gca	gag	agc	ctg	tcg	gta	539
	Ala	His	Met	Phe	Val	Lys	Leu	Lys	Val	Ile	Ala	Glu	Ser	Leu	Ser	Val	
				120					125					130			
	att	ctt	aag	acc	gtt	ctg	aca	gct	ttt	ctc	gtg	ctg	tgg	ttg	cct	cac	587
	Ile	Leu	Lys	Thr	Val	Leu	Thr	Ala	Phe	Leu	Val	Leu	Trp	Leu	Pro	His	
				135				140					145				
25	tgg	gga	ttg	tac	att	ttc	tct	ttg	gcc	cag	ctt	ttc	tat	acc	aca	gtt	635
	Trp	Gly	Leu	Tyr	Ile	Phe	Ser	Leu	Ala	Gln	Leu	Phe	Tyr	Thr	Thr	Val	
				150			155					160					
	ctg	gtg	ctc	tgc	tat	gtt	att	tat	ttc	aca	aag	tta	ctg	ggg	tcc	cca	683
	Leu	Val	Leu	Cys	Tyr	Val	Ile	Tyr	Phe	Thr	Lys	Leu	Leu	Gly	Ser	Pro	
30	165					170						175				180	
	gaa	tca	acc	aag	ctt	caa	act	ctt	cct	gtc	tcc	aga	ata	aca	gat	ctg	731
	Glu	Ser	Thr	Lys	Leu	Gln	Thr	Leu	Pro	Val	Ser	Arg	Ile	Thr	Asp	Leu	
					185					190					195		
35	tta	ccc	aat	att	aca	aga	aat	gga	gcg	ttt	ata	aac	tgg	aaa	gag	gct	779
	Leu	Pro	Asn	Ile	Thr	Arg	Asn	Gly	Ala	Phe	Ile	Asn	Trp	Lys	Glu	Ala	
				200					205					210			
	aaa	ctg	act	tgg	agt	ttt	ttc	aaa	cag	tct	ttc	ttg	aaa	cag	att	ttg	827
	Lys	Leu	Thr	Trp	Ser	Phe	Phe	Lys	Gln	Ser	Phe	Leu	Lys	Gln	Ile	Leu	
				215				220					225				
40	aca	gaa	ggc	gag	cga	tat	gtg	atg	aca	ttt	ttg	aat	gta	ttg	aac	ttt	875
	Thr	Glu	Gly	Glu	Arg	Tyr	Val	Met	Thr	Phe	Leu	Asn	Val	Leu	Asn	Phe	
				230			235					240					
	ggg	gat	cag	ggg	gtg	tat	gat	ata	gtg	aat	aat	ctt	ggc	tcc	ctt	gtg	923
	Gly	Asp	Gln	Gly	Val	Tyr	Asp	Ile	Val	Asn	Asn	Leu	Gly	Ser	Leu	Val	
45	245					250						255				260	
	gcc	aga	tta	att	ttc	cag	cca	ata	gag	gaa	agt	ttt	tat	ata	ttt	ttt	971
	Ala	Arg	Leu	Ile	Phe	Gln	Pro	Ile	Glu	Glu	Ser	Phe	Tyr	Ile	Phe	Phe	
					265					270						275	
50	gct	aag	gtg	ctg	gag	agg	gga	aag	gat	gcc	aca	ctt	cag	aag	cag	gag	1019
	Ala	Lys	Val	Leu	Glu	Arg	Gly	Lys	Asp	Ala	Thr	Leu	Gln	Lys	Gln	Glu	
				280					285						290		
	gac	gtt	gct	gtg	gct	gct	gca	gtc	ttg	gag	tcc	ctg	ctc	aag	ctg	gcc	1067
	Asp	Val	Ala	Val	Ala	Ala	Ala	Val	Leu	Glu	Ser	Leu	Lys	Lys	Leu	Ala	
				295					300					305			
55	ctg	ctg	gcc	ggc	ctg	acc	atc	act	gtt	ttt	ggc	ttt	gcc	tat	tct	cag	1115
	Leu	Leu	Ala	Gly	Leu	Thr	Ile	Thr	Val	Phe	Gly	Phe	Ala	Tyr	Ser	Gln	
				310				315				320					
	ctg	gct	ctg	gat	atc	tac	gga	ggg	acc	atg	ctt	agc	tca	gga	tcc	ggg	1163
	Leu	Ala	Leu	Asp	Ile	Tyr	Gly	Gly	Thr	Met	Leu	Ser	Ser	Gly	Ser	Gly	
60	325					330						335				340	
	cct	gtt	ttg	ctg	cgt	tcc	tac	tgt	ctc	tat	gtt	ctc	ctg	ctt	gcc	atc	1211
	Pro	Val	Leu	Leu	Arg	Ser	Tyr	Cys	Leu	Tyr	Val	Leu	Leu	Leu	Ala	Ile	
					345					350						355	

	aat gga gtg aca gag tgt ttc aca ttt gct gcc atg agc aaa gag gag	1259
	Asn Gly Val Thr Glu Cys Phe Thr Phe Ala Ala Met Ser Lys Glu Glu	
	360 365 370	
5	gtc gac agg tac aat ttt gtg atg ctg gcc ctg tcc tcc tca ttc ctg	1307
	Val Asp Arg Tyr Asn Phe Val Met Leu Ala Leu Ser Ser Ser Phe Leu	
	375 380 385	
	gtg tta tcc tat ctc ttg acc cgt tgg tgt ggc agc gtg ggc ttc atc	1355
	Val Leu Ser Tyr Leu Leu Thr Arg Trp Cys Gly Ser Val Gly Phe Ile	
	390 395 400	
10	ttg gcc aac tgc ttt aac atg ggc att cgg atc acg cag agc ctt tgc	1403
	Leu Ala Asn Cys Phe Asn Met Gly Ile Arg Ile Thr Gln Ser Leu Cys	
	405 410 415 420	
	ttc atc cac cgc tac tac cga agg agc ccc cac agg ccc ctg gct ggc	1451
	Phe Ile His Arg Tyr Tyr Arg Arg Ser Pro His Arg Pro Leu Ala Gly	
15	425 430 435	
	ctg cac cta tcg cca gtc ctg ctc ggg aca ttt gcc ctc agt ggt ggg	1499
	Leu His Leu Ser Pro Val Leu Leu Gly Thr Phe Ala Leu Ser Gly Gly	
	440 445 450	
20	gtt act gct gtt tcg gag gta ttc ctc tgc tgt gag cag ggc tgg cca	1547
	Val Thr Ala Val Ser Glu Val Phe Leu Cys Cys Glu Gln Gly Trp Pro	
	455 460 465	
	gcc aga ctg gca cac att gct gtg ggg gcc ttc tgt ctg gga gca act	1595
	Ala Arg Leu Ala His Ile Ala Val Gly Ala Phe Cys Leu Gly Ala Thr	
	470 475 480	
25	ctc ggg aca gca ttc ctc aca gag acc aag ctg atc cat ttc ctc agg	1643
	Leu Gly Thr Ala Phe Leu Thr Glu Thr Lys Leu Ile His Phe Leu Arg	
	485 490 495 500	
	act cag tta ggt gtg ccc aga cgc act gac aaa atg aca tgacttcagg	1692
	Thr Gln Leu Gly Val Pro Arg Arg Thr Asp Lys Met Thr	
30	505 510	
	gaagcctgga caccgaggc acctggacca gctatgggta gttctgtggg tggaacacat	1752
	tctgtgtaag agccccactg agggctctgc agcggagtga cagcaacccc agagatgagg	1812
	caccagagag tgccactgca tgagacacct gtgaccattc gaagtctgaa atgcgggggg	1872
	ggagtttcat ttttaagtga agaccaaaag ccctttaaaa ataatagttt tttatcaaaa	1932
35	aaaaaaaaaaa aaa	1945
	<210> 38	
	<211> 1330	
	<212> DNA	
40	<213> Homo sapiens	
	<220>	
	<221> CDS	
	<222> 198..998	
45	<220>	
	<221> sig_peptide	
	<222> 198..269	
	<223> Von Heijne matrix	
50	score 9.08017839002281	
	seq LLLGPGLLATVRA/EC	
	<400> 38	
	agaaatcagc cctttgcaga gggcgagag ggccctggaaa cctctgggac cttttcccag	60
55	gaactgttta tggtttcccc ctaggctctag gagacgtaga tgcataagggtg gattggatac	120
	atcgatggta gctataagag tcgtgtctga acccggttt tccaattggc ctgctccatc	180
	cgaacagcgt caactcc atg gcg cgg ttc ctg aca ctt tgc act tgg ctg	230
	Met Ala Arg Phe Leu Thr Leu Cys Thr Trp Leu	
	-20 -15	
60	ctg ttg ctc ggc ccc ggg ctc ctg gcg acc gtg cgg gcc gaa tgc agc	278
	Leu Leu Leu Gly Pro Gly Leu Leu Ala Thr Val Arg Ala Glu Cys Ser	
	-10 -5 1	
	cag gat tgc gcg acg tgc agc tac cgc cta gtg cgc ccg gcc gac atc	326

	Gln	Asp	Cys	Ala	Thr	Cys	Ser	Tyr	Arg	Leu	Val	Arg	Pro	Ala	Asp	Ile	
	5						10					15					
	aac	ttc	ctg	gct	tgc	gta	atg	gaa	tgt	gaa	ggc	aaa	ctg	cct	tct	ctg	374
	Asn	Phe	Leu	Ala	Cys	Val	Met	Glu	Cys	Glu	Gly	Lys	Leu	Pro	Ser	Leu	
5	20					25				30					35		
	aaa	att	tgg	gaa	acc	tgc	aag	gag	ctc	ctg	cag	ctg	tcc	aaa	cca	gat	422
	Lys	Ile	Trp	Glu	Thr	Cys	Lys	Glu	Leu	Leu	Gln	Leu	Ser	Lys	Pro	Asp	
					40					45					50		
	ctt	cct	caa	gat	ggc	acc	agc	acc	ctc	aga	gaa	aat	agc	aaa	ccg	gaa	470
10	Leu	Pro	Gln	Asp	Gly	Thr	Ser	Thr	Leu	Arg	Glu	Asn	Ser	Lys	Pro	Glu	
					55					60					65		
	gaa	agc	cat	ttg	cta	gcc	aaa	agg	tat	ggg	ggc	ttc	atg	aaa	agg	tat	518
	Glu	Ser	His	Leu	Leu	Ala	Lys	Arg	Tyr	Gly	Gly	Phe	Met	Lys	Arg	Tyr	
					70				75					80			
15	gga	ggc	ttc	atg	aag	aaa	atg	gat	gag	ctt	tat	ccc	atg	gag	cca	gaa	566
	Gly	Gly	Phe	Met	Lys	Lys	Met	Asp	Glu	Leu	Tyr	Pro	Met	Glu	Pro	Glu	
					85				90					95			
	gaa	gag	gcc	aat	gga	agt	gag	atc	ctc	gcc	aag	cgg	tat	ggg	ggc	ttc	614
	Glu	Glu	Ala	Asn	Gly	Ser	Glu	Ile	Leu	Ala	Lys	Arg	Tyr	Gly	Gly	Phe	
20						105											
	atg	aag	aag	gat	gca	gag	gag	gac	gac	tcg	ctg	gcc	aat	tcc	tca	gac	662
	Met	Lys	Lys	Asp	Ala	Glu	Glu	Asp	Asp	Ser	Leu	Ala	Asn	Ser	Ser	Asp	
					120					125					130		
	ctg	cta	aaa	gag	ctt	ctg	gaa	aca	ggg	gac	aac	cga	gag	cgt	agc	cac	710
25	Leu	Leu	Lys	Glu	Leu	Leu	Glu	Thr	Gly	Asp	Asn	Arg	Glu	Arg	Ser	His	
					135				140					145			
	cac	cag	gat	ggc	agt	gat	aat	gag	gaa	gaa	gtg	agc	aag	aga	tat	ggg	758
	His	Gln	Asp	Gly	Ser	Asp	Asn	Glu	Glu	Glu	Val	Ser	Lys	Arg	Tyr	Gly	
					150				155					160			
30	ggc	ttc	atg	aga	ggc	tta	aag	aga	agc	ccc	caa	ctg	gaa	gat	gaa	gcc	806
	Gly	Phe	Met	Arg	Gly	Leu	Lys	Arg	Ser	Pro	Gln	Leu	Glu	Asp	Glu	Ala	
					165				170					175			
	aaa	gag	ctg	cag	aag	cga	tat	ggg	ggc	ttc	atg	aga	aga	gta	ggc	cgc	854
	Lys	Glu	Leu	Gln	Lys	Arg	Tyr	Gly	Gly	Phe	Met	Arg	Arg	Val	Gly	Arg	
35	180					185				190					195		
	cca	gag	tgg	tgg	atg	gac	tac	cag	aaa	cgg	tat	gga	ggc	ttc	ctg	aag	902
	Pro	Glu	Trp	Trp	Met	Asp	Tyr	Gln	Lys	Arg	Tyr	Gly	Gly	Phe	Leu	Lys	
					200					205					210		
	cgc	ttt	gcc	gag	gct	ctg	ccc	tcc	gac	gaa	gaa	ggc	gaa	agt	tac	tcc	950
40	Arg	Phe	Ala	Glu	Ala	Leu	Pro	Ser	Asp	Glu	Glu	Gly	Glu	Ser	Tyr	Ser	
					215				220					225			
	aaa	gaa	gtt	cct	gaa	atg	gaa	aaa	aga	tac	gga	gga	ttt	atg	aga	ttt	998
	Lys	Glu	Val	Pro	Glu	Met	Glu	Lys	Arg	Tyr	Gly	Gly	Phe	Met	Arg	Phe	
					230				235					240			
45	taatat	ttttt	cccactag	tg	gccccagg	cc	cagcaag	cc	tcctccat	c	ctccagt	ggg					1058
	aaactg	tttga	ttgtcat	gtg	ttgcttg	cct	tgtatag	ttg	acttcatt	gt							1118
	ctggata	act	atacaac	ctg	aaaactg	tca	tttcagg	ttc	tgtgct	ctt	ttggagt	ctt					1178
	taagct	cagt	attagt	ctat	tcagct	atc	tcgtttt	cat	gctaaa	atag	ttttt	gttat					1238
	cttgt	ctctt	attttt	gaca	aacat	caata	aatgct	tact	tgtat	ataga	gataa	taaac					1298
50	ctattac	cccc	aagtgc	aaaaa	aaaaaaaa	aa											1330

<210> 39
 <211> 2124
 <212> DNA
 55 <213> Homo sapiens

<220>
 <221> CDS
 <222> 505..1590
 60
 <220>
 <221> sig_peptide
 <222> 505..624

[illegible]

```

210          215          220          225
att ggg aac tgt gtc tgt ctc ctg gtc cta agc tca gca ctt cct gtc      1347
Ile Gly Asn Cys Val Cys Leu Leu Val Leu Ser Ser Ala Leu Pro Val

          230          235          240
5 ttc tct cga acc ctg ggg ctc act cgc ttt gac ctg ctg ggt gac ttt      1395
Phe Ser Arg Thr Leu Gly Leu Thr Arg Phe Asp Leu Leu Gly Asp Phe

          245          250          255
gga cgc ttc aac tgg ctg ggc aat ttc tac att gtg ttc ctc tac aac      1443
Gly Arg Phe Asn Trp Leu Gly Asn Phe Tyr Ile Val Phe Leu Tyr Asn

10          260          265          270
gca gcc ttt gca ggc ctc acc aca ctc tat ctg gtg aag acc ttc act      1491
Ala Ala Phe Ala Gly Leu Thr Thr Leu Tyr Leu Val Lys Thr Phe Thr

          275          280          285
gca gct gtg cgg gca gag ctg atc cgg gcc ttt ggg ctg gac aga ctg      1539
15 Ala Ala Val Arg Ala Glu Leu Ile Arg Ala Phe Gly Leu Asp Arg Leu

          290          295          300          305
ccg ctg ccc gtc tcc ggt ttc ccc cag gca tct agg aag acc cag cac      1587
Pro Leu Pro Val Ser Gly Phe Pro Gln Ala Ser Arg Lys Thr Gln His

          310          315          320
20 cag tgacctccag ctgggggtgg gaagaaaaaa actggacact gccatctgct      1640
Gln

gcctaggcct ggaggggaagc ccaaggctac ttggacctca ggacctggaa tctgagaggg      1700
tgggtggcag aggggagcag agccatctgc actattgcat aatctgagcc agagtttggg      1760
accaggacct cctgcttttc cataactaac tgtggcctca gcatggggta gggctgggtg      1820
25 actgggtcta gcccctgatc ccaaactctgt ttacacatca atctgcctca ctgctgttct      1880
gggccatccc catagccatg tttacatgat ttgatgtgca ataggggtggg gtaggggcag      1940
ggaaaggact gggccagggc aggctcggga gatagattgt ctcccttgcc tctggcccag      2000
cagagcctaa gcaactgtgct atcctggagg ggctttggac cacctgaaag accaagggga      2060
tagggaggag gaggcttcag ccatacagcaa taaagttagt cccaggcaaa aaaaaaaaaa      2120
30 aaaa      2124

<210> 40
<211> 1159
<212> DNA
35 <213> Homo sapiens

<220>
<221> CDS
<222> 84..326

40
<220>
<221> sig_peptide
<222> 84..146
<223> Von Heijne matrix
45 score 6.39000252120129
seq LGLSVLLTAATVA/GV

<400> 40
agtacaggcg gcggtgcgca ctctgcggcg gcctctgcgc ctgggcggg cgaggagagag      60
50 aggccgcggc cgccagcgtg ggg atg tct agg agc tcg aag gtg gtg ctg ggc      113
Met Ser Arg Ser Ser Lys Val Val Leu Gly

-20
ctc tcg gtg ctg ctg acg gcg gcc aca gtg gcc ggc gta cat gtg aag      161
Leu Ser Val Leu Leu Thr Ala Ala Thr Val Ala Gly Val His Val Lys

55 -10 -5 1 5
cag cag tgg gac cag cag agg ctt cgt gac gga gtt atc aga gac att      209
Gln Gln Trp Asp Gln Gln Arg Leu Arg Asp Gly Val Ile Arg Asp Ile

10 15 20
gag agg caa att cgg aaa aaa gaa aac att cgt ctt ttg gga gaa cag      257
60 Glu Arg Gln Ile Arg Lys Lys Glu Asn Ile Arg Leu Leu Gly Glu Gln

25 30 35
att att ttg act gag caa ctt gaa gca gaa aga gag aag atg tta ttg      305
Ile Ile Leu Thr Glu Gln Leu Glu Ala Glu Arg Glu Lys Met Leu Leu

```

	40	45	50	
	gca aaa gga tct caa aaa tca tgacttgaat gtgaaatatac tgttggacag			356
	Ala Lys Gly Ser Gln Lys Ser			
	55	60		
5	acaacacgag tttgtgtgtg tgtgttgatg gagagtagct tagtagtatac ttcattctttt			416
	tttttgggtca ctgtcctttt aaacttgatc aaataaagga cagtgggtca tataagttac			476
	tgctttcagg gtcccttata tctgaataaa ggagtgtggg cagacacttt ttggaagagt			536
	ctgtctgggt gatcctggta gaagcccat tagggctact gtccagtgtc tagggttgtt			596
	actgagaagc actgccgagc ttgtgagaag gaagggatgg atagtagcat ccacctgagt			656
10	agtctgatca gtcggcatga tgacgaagcc acgagaacat cgacctcaga aggactggag			716
	gaaggtgaaa gtggagggag agacgctcct gatcgtcgaa tyccgaggat caggkcatca			776
	gtggacttat cgcacgacca gagtggggat tccctcaaca gtgatgaagg agacgtgtct			836
	tggatggagg agcagctgtc ctacttctgt gacaagtgcc aaaaatggat accagccagt			896
	aaggagcttc tcaattcctt tgatttgtca attcctgtgt gaaggtttgt ttttccaacc			956
15	tgtgaaagaa acgtgaatgt aaaagagacc taaataaaag gataattata tttattctct			1016
	agttgatcag ctataaattt atataaaaca taggcatgtt tgtactaatg aaacgtactg			1076
	tcaacctcta tcacattgtt aaattaacac ttttgggtgt aactcaataa aattgagaaa			1136
	attgcaaaaa aaaaaaaaaa aaa			1159
20	<210> 41			
	<211> 1953			
	<212> DNA			
	<213> Homo sapiens			
25	<220>			
	<221> CDS			
	<222> 56..1678			
	<220>			
30	<221> sig_peptide			
	<222> 56..139			
	<223> Von Heijne matrix			
	score 3.75144398608723			
	seq SGLLLQVLFRLIT/FV			
35	<400> 41			
	agactaaacg gagtggcggc ggcatttcct ggtgtctgag cctggcgcgagg aggct atg			58
			Met	
	ggc agc cag gag gtg ctg ggc cac gcg gcc cgg ctg gcc tcc tcc ggt			106
40	Gly Ser Gln Glu Val Leu Gly His Ala Ala Arg Leu Ala Ser Ser Gly			
	-25 -20 -15			
	ctc ctc ctg cag gtg ttg ttt cgg ttg atc acc ttt gtc ttg aat gca			154
	Leu Leu Leu Gln Val Leu Phe Arg Leu Ile Thr Phe Val Leu Asn Ala			
	-10 -5 1 5			
45	ttt att ctt cgc ttc ctg tca aag gaa atc gtt ggc gta gta aat gta			202
	Phe Ile Leu Arg Phe Leu Ser Lys Glu Ile Val Gly Val Val Asn Val			
	10 15 20			
	aga cta acg ctg ctt tac tca acc acc ctc ttc ctg gcc aga gag gcc			250
	Arg Leu Thr Leu Leu Tyr Ser Thr Thr Leu Phe Leu Ala Arg Glu Ala			
50	25 30 35			
	ttc cgc aga gca tgt ctc agt ggg ggc acc cag cga gac tgg agc cag			298
	Phe Arg Arg Ala Cys Leu Ser Gly Gly Thr Gln Arg Asp Trp Ser Gln			
	40 45 50			
	acc ctc aac ctg ctg tgg cta aca gtc ccc ctg ggt gtg ttt tgg tcc			346
55	Thr Leu Asn Leu Leu Trp Leu Thr Val Pro Leu Gly Val Phe Trp Ser			
	55 60 65			
	tta ttc ctg ggc tgg atc tgg ttg cag ctg ctt gaa gtg cct gat cct			394
	Leu Phe Leu Gly Trp Ile Trp Leu Gln Leu Leu Glu Val Pro Asp Pro			
	70 75 80 85			
60	aat gtt gtc cct cac tat gca act gga gtg gtg ctg ttt ggt ctc tcg			442
	Asn Val Val Pro His Tyr Ala Thr Gly Val Val Leu Phe Gly Leu Ser			
	90 95 100			
	gca gtg gtg gag ctt cta gga gag ccc ttt tgg gtc ttg gca caa gca			490

	Ala	Val	Val	Glu	Leu	Leu	Gly	Glu	Pro	Phe	Trp	Val	Leu	Ala	Gln	Ala	
				105					110					115			
	cat	atg	ttt	gtg	aag	ctc	aag	gtg	att	gca	gag	agc	ctg	tcg	gta	att	538
	His	Met	Phe	Val	Lys	Leu	Lys	Val	Ile	Ala	Glu	Ser	Leu	Ser	Val	Ile	
5			120					125					130				
	ctt	aag	agc	gtt	ctg	aca	gct	ttt	ctc	gtg	ctg	tgg	ttg	cct	cac	tgg	586
	Leu	Lys	Ser	Val	Leu	Thr	Ala	Phe	Leu	Val	Leu	Trp	Leu	Pro	His	Trp	
			135				140						145				
	gga	ttg	tac	att	ttc	tct	ttg	gcc	cag	ctt	ttc	tat	acc	aca	gtt	ctg	634
10	Gly	Leu	Tyr	Ile	Phe	Ser	Leu	Ala	Gln	Leu	Phe	Tyr	Thr	Thr	Val	Leu	
	150					155					160					165	
	gtg	ctc	tgc	tat	gtt	att	tat	ttc	aca	aag	tta	ctg	ggg	tcc	cca	gaa	682
	Val	Leu	Cys	Tyr	Val	Ile	Tyr	Phe	Thr	Lys	Leu	Leu	Gly	Ser	Pro	Glu	
				170						175					180		
15	tca	acc	aag	ctt	caa	act	ctt	cct	gtc	tcc	aga	ata	aca	gat	ctg	tta	730
	Ser	Thr	Lys	Leu	Gln	Thr	Leu	Pro	Val	Ser	Arg	Ile	Thr	Asp	Leu	Leu	
			185						190					195			
	ccc	aat	att	aca	aga	aat	gga	gcg	ttt	ata	aac	tgg	aaa	gag	gct	aaa	778
20	Pro	Asn	Ile	Thr	Arg	Asn	Gly	Ala	Phe	Ile	Asn	Trp	Lys	Glu	Ala	Lys	
			200					205					210				
	ctg	act	tgg	agt	ttt	ttc	aaa	cag	tct	ttc	ttg	aaa	cag	att	ttg	aca	826
	Leu	Thr	Trp	Ser	Phe	Phe	Lys	Gln	Ser	Phe	Leu	Lys	Gln	Ile	Leu	Thr	
			215				220						225				
25	gaa	ggc	gag	cga	tat	gtg	atg	aca	ttt	ttg	aat	gta	ttg	aac	ttt	ggg	874
	Glu	Gly	Glu	Arg	Tyr	Val	Met	Thr	Phe	Leu	Asn	Val	Leu	Asn	Phe	Gly	
	230					235					240				245		
	gat	cag	ggg	gtg	tat	gat	ata	gtg	aat	aat	ctt	ggc	tcc	ctt	gtg	gcc	922
	Asp	Gln	Gly	Val	Tyr	Asp	Ile	Val	Asn	Asn	Leu	Gly	Ser	Leu	Val	Ala	
				250					255					260			
30	aga	tta	att	ttc	cag	cca	ata	gag	gaa	agt	ttt	tat	ata	ttt	ttt	gct	970
	Arg	Leu	Ile	Phe	Gln	Pro	Ile	Glu	Glu	Ser	Phe	Tyr	Ile	Phe	Phe	Ala	
				265					270					275			
	aag	gtg	ctg	gag	agg	gga	aag	gat	gcc	aca	ctt	cag	aag	cag	gag	gac	1018
	Lys	Val	Leu	Glu	Arg	Gly	Lys	Asp	Ala	Thr	Leu	Gln	Lys	Gln	Glu	Asp	
35			280					285					290				
	gtt	gct	gtg	gct	gct	gca	gtc	ttg	gag	tcc	ctg	ctc	aag	ctg	gcc	ctg	1066
	Val	Ala	Val	Ala	Ala	Ala	Val	Leu	Glu	Ser	Leu	Lys	Leu	Ala	Leu		
			295				300						305				
	ctg	gcc	ggc	ctg	acc	atc	act	gtt	ttt	ggc	ttt	gcc	tat	tct	cag	ctg	1114
40	Leu	Ala	Gly	Leu	Thr	Ile	Thr	Val	Phe	Gly	Phe	Ala	Tyr	Ser	Gln	Leu	
	310					315					320				325		
	gct	ctg	gat	atc	aac	gga	ggg	acc	atg	ctt	agc	tca	gga	tcc	ggg	cct	1162
	Ala	Leu	Asp	Ile	Asn	Gly	Gly	Thr	Met	Leu	Ser	Ser	Gly	Ser	Gly	Pro	
				330					335						340		
45	gtt	ttg	ctg	cgt	tcc	tac	tgt	ctc	tat	gtt	ctc	ctg	ctt	gcc	atc	aat	1210
	Val	Leu	Leu	Arg	Ser	Tyr	Cys	Leu	Tyr	Val	Leu	Leu	Leu	Ala	Ile	Asn	
				345					350					355			
	gga	gtg	aca	gag	tgt	ttc	aca	ttt	gct	gcc	atg	agc	aaa	gag	gag	gtc	1258
	Gly	Val	Thr	Glu	Cys	Phe	Thr	Phe	Ala	Ala	Met	Ser	Lys	Glu	Glu	Val	
50			360					365					370				
	gac	agg	tac	aat	ttt	gtg	atg	ctg	gcc	ctg	tcc	tcc	tca	ttc	ctg	gtg	1306
	Asp	Arg	Tyr	Asn	Phe	Val	Met	Leu	Ala	Leu	Ser	Ser	Ser	Phe	Leu	Val	
			375				380					385					
	tta	tcc	tat	ctc	ttg	acc	cgt	tgg	tgt	ggc	agc	gtg	ggc	ttc	atc	ttg	1354
55	Leu	Ser	Tyr	Leu	Leu	Thr	Arg	Trp	Cys	Gly	Ser	Val	Gly	Phe	Ile	Leu	
	390					395					400				405		
	gcc	aac	tgc	ttt	aac	atg	ggc	att	cgg	atc	acg	cag	agc	ctt	tgc	ttc	1402
	Ala	Asn	Cys	Phe	Asn	Met	Gly	Ile	Arg	Ile	Thr	Gln	Ser	Leu	Cys	Phe	
				410					415						420		
60	atc	cac	cgc	tac	tac	cga	agg	agc	ccc	cac	agg	ccc	ctg	gct	ggc	ctg	1450
	Ile	His	Arg	Tyr	Tyr	Arg	Arg	Ser	Pro	His	Arg	Pro	Leu	Ala	Gly	Leu	
				425					430					435			
	cac	cta	tcg	cca	gtc	ctg	ctc	ggg	aca	ttt	gcc	ctc	agt	ggg	ggg	gtt	1498

His Leu Ser Pro Val Leu Leu Gly Thr Phe Ala Leu Ser Gly Gly Val
 440 445 450
 act gct gtt tcg gag gta ttc ctc tgc tgt gag cag ggc tgg cca gcc 1546
 Thr Ala Val Ser Glu Val Phe Leu Cys Cys Glu Gln Gly Trp Pro Ala
 5 455 460 465
 aga ctg gca cac att gct gtg ggg gcc ttc tgt ctg gga gca act ctc 1594
 Arg Leu Ala His Ile Ala Val Gly Ala Phe Cys Leu Gly Ala Thr Leu
 470 475 480 485
 ggg aca gca ttc ctc aca gag acc aag ctg atc cat ttc ctc agg act 1642
 10 Gly Thr Ala Phe Leu Thr Glu Thr Lys Leu Ile His Phe Leu Arg Thr
 490 495 500
 cag tta ggt gtg ccc aga cgc act gac aaa atg acg tgacttcagg 1688
 Gln Leu Gly Val Pro Arg Arg Thr Asp Lys Met Thr
 505 510
 15 gaagcctgga caccgaggc acctggacca gctatgggta gttctgtggg tggaacacat 1748
 tctgtgtaag agcccactg agggctctgc agcggagtga cagcaacccc agagatgagg 1808
 caccagagag tgccactgca tgagacacct gtgaccattc gaagtctgaa atgcgggggg 1868
 ggagtttcat ttttaagtga agaccaaag ccctttaaaa ataatagttt tttatcattt 1928
 tatagtga aa aaaaaaaa aaaaa 1953
 20
 <210> 42
 <211> 1688
 <212> DNA
 <213> Homo sapiens
 25
 <220>
 <221> CDS
 <222> 119..1522
 30 <220>
 <221> sig_peptide
 <222> 119..181
 <223> Von Heijne matrix
 score 11.6921972463885
 35 seq LLLCLALSGAAET/KP
 <400> 42
 aaaaggctgc aggetgccag gtgtgcttgg agagccccct tcttccgccg ggcctcgcaa 60
 gcagcgtagg actgtggaga agggcggtgg gcaaggaggg aactcgagag cagcctcc 118
 40 atg ggc aca cag gag ggc tgg tgc ctg ctg ctc tgc ctg gct cta tct 166
 Met Gly Thr Gln Glu Gly Trp Cys Leu Leu Leu Cys Leu Ala Leu Ser
 -20 -15 -10
 gga gca gca gaa acc aag ccc cac cca gca gag ggg cag tgg cgg gca 214
 Gly Ala Ala Glu Thr Lys Pro His Pro Ala Glu Gly Gln Trp Arg Ala
 45 -5 1 5 10
 gtg gac gtg gtc cta gac tgt ttc ctg gtg aag gac ggt gcg cac cgt 262
 Val Asp Val Val Leu Asp Cys Phe Leu Val Lys Asp Gly Ala His Arg
 15 20 25
 gga gct ctc gcc agc agt gag gac agg gca agg gcc tcc ctt gtg ctg 310
 50 Gly Ala Leu Ala Ser Ser Glu Asp Arg Ala Arg Ala Ser Leu Val Leu
 30 35 40
 aag cag gtg cca gtg ctg gac gat ggc tcc ctg gag gac ttc acc gat 358
 Lys Gln Val Pro Val Leu Asp Gly Ser Leu Glu Asp Phe Thr Asp
 45 50 55
 55 ttc caa ggg ggc aca ctg gcc caa gat gac cca cct att atc ttt gag 406
 Phe Gln Gly Gly Thr Leu Ala Gln Asp Asp Pro Pro Ile Ile Phe Glu
 60 65 70 75
 gcc tca gtg gac ctg gtc cag att ccc cag gcc gag gcc ttg ctc cat 454
 Ala Ser Val Asp Leu Val Gln Ile Pro Gln Ala Glu Ala Leu Leu His
 60 80 85 90
 gct gac tgc agt ggg aag gag gtg acc tgt gag atc tcc cgc tac ttt 502
 Ala Asp Cys Ser Gly Lys Glu Val Thr Cys Glu Ile Ser Arg Tyr Phe
 95 100 105

	ctc	cag	atg	aca	gag	acc	act	gtt	aag	aca	gca	gct	tgg	ttc	atg	gcc	550
	Leu	Gln	Met	Thr	Glu	Thr	Thr	Val	Lys	Thr	Ala	Ala	Trp	Phe	Met	Ala	
			110					115					120				
	aac	gtg	cag	gtc	tct	gga	ggg	gga	cct	agc	atc	tcc	ttg	gtg	atg	aag	598
5	Asn	Val	Gln	Val	Ser	Gly	Gly	Gly	Pro	Ser	Ile	Ser	Leu	Val	Met	Lys	
			125					130					135				
	act	ccc	agg	gtc	gcc	aag	aat	gag	gtg	ctc	tgg	cac	cca	acg	ctg	aac	646
	Thr	Pro	Arg	Val	Ala	Lys	Asn	Glu	Val	Leu	Trp	His	Pro	Thr	Leu	Asn	
						145					150					155	
10	ttg	cca	ctg	agc	ccc	cag	ggg	act	gtg	cga	act	gca	gtg	gag	ttc	cag	694
	Leu	Pro	Leu	Ser	Pro	Gln	Gly	Thr	Val	Arg	Thr	Ala	Val	Glu	Phe	Gln	
					160						165					170	
	gtg	atg	aca	cag	acc	caa	tcc	ctg	agc	ttc	ctg	ctg	ggg	tcc	tca	gcc	742
	Val	Met	Thr	Gln	Thr	Gln	Ser	Leu	Ser	Phe	Leu	Leu	Gly	Ser	Ser	Ala	
15					175					180						185	
	tcc	ttg	gac	tgt	ggc	ttc	tcc	atg	gca	ccg	ggc	ttg	gac	ctc	atc	agt	790
	Ser	Leu	Asp	Cys	Gly	Phe	Ser	Met	Ala	Pro	Gly	Leu	Asp	Leu	Ile	Ser	
			190					195					200				
	gtg	gag	tgg	cga	ctg	cag	cac	aag	ggc	agg	ggt	cag	ttg	gtg	tac	agc	838
20	Val	Glu	Trp	Arg	Leu	Gln	His	Lys	Gly	Arg	Gly	Gln	Leu	Val	Tyr	Ser	
			205					210					215				
	tgg	acc	gca	ggg	cag	ggg	cag	gct	gtg	cgg	aag	ggc	gct	acc	ctg	gag	886
	Trp	Thr	Ala	Gly	Gln	Gly	Gln	Ala	Val	Arg	Lys	Gly	Ala	Thr	Leu	Glu	
			220			225					230					235	
25	cct	gca	caa	ctg	ggc	atg	gcc	agg	gat	gcc	tcc	ctc	acc	ctg	ccc	ggc	934
	Pro	Ala	Gln	Leu	Gly	Met	Ala	Arg	Asp	Ala	Ser	Leu	Thr	Leu	Pro	Gly	
					240					245					250		
	ctc	act	ata	cag	gac	gag	ggg	acc	tac	att	tgc	cag	atc	acc	acc	tct	982
	Leu	Thr	Ile	Gln	Asp	Glu	Gly	Thr	Tyr	Ile	Cys	Gln	Ile	Thr	Thr	Ser	
30					255					260						265	
	ctg	tac	cga	gct	cag	cag	atc	atc	cag	ctc	aac	atc	caa	gct	tcc	cct	1030
	Leu	Tyr	Arg	Ala	Gln	Gln	Ile	Ile	Gln	Leu	Asn	Ile	Gln	Ala	Ser	Pro	
			270					275								280	
	aaa	gta	cga	ctg	agc	ttg	gca	aac	gaa	gct	ctg	ctg	ccc	acc	ctc	atc	1078
35	Lys	Val	Arg	Leu	Ser	Leu	Ala	Asn	Glu	Ala	Leu	Leu	Pro	Thr	Leu	Ile	
			285					290					295				
	tgc	gac	att	gct	ggc	tat	tac	cct	ctg	gat	gtg	gtg	gtg	acg	tgg	acc	1126
	Cys	Asp	Ile	Ala	Gly	Tyr	Tyr	Pro	Leu	Asp	Val	Val	Val	Thr	Trp	Thr	
			300			305					310					315	
40	cga	gag	gag	ctg	ggt	gga	tcc	cca	gcc	caa	gtc	tct	ggt	gcc	tcc	ttc	1174
	Arg	Glu	Glu	Leu	Gly	Gly	Ser	Pro	Ala	Gln	Val	Ser	Gly	Ala	Ser	Phe	
					320					325						330	
	tcc	agc	ctc	agg	caa	agc	gtg	gca	ggc	acc	tac	agc	atc	tcc	tcc	tct	1222
	Ser	Ser	Leu	Arg	Gln	Ser	Val	Ala	Gly	Thr	Tyr	Ser	Ile	Ser	Ser	Ser	
45					335					340						345	
	ctc	acc	gca	gaa	cct	ggc	tct	gca	ggt	gcc	act	tac	acc	tgc	cag	gtc	1270
	Leu	Thr	Ala	Glu	Pro	Gly	Ser	Ala	Gly	Ala	Thr	Tyr	Thr	Cys	Gln	Val	
			350					355								360	
	aca	cac	atc	tct	ctg	gag	gag	ccc	ctt	ggg	gcc	agc	acc	cag	gtt	gtc	1318
50	Thr	His	Ile	Ser	Leu	Glu	Glu	Pro	Leu	Gly	Ala	Ser	Thr	Gln	Val	Val	
			365					370								375	
	cca	cca	gag	cgg	aga	aca	gcc	ttg	gga	gtc	atc	ttt	gcc	agc	agt	ctc	1366
	Pro	Pro	Glu	Arg	Arg	Thr	Ala	Leu	Gly	Val	Ile	Phe	Ala	Ser	Ser	Leu	
			380			385						390				395	
55	ttc	ctt	ctt	gca	ctg	atg	ttc	ctg	ggg	ctt	cag	aga	cgg	caa	gca	cct	1414
	Phe	Leu	Leu	Ala	Leu	Met	Phe	Leu	Gly	Leu	Gln	Arg	Arg	Gln	Ala	Pro	
					400					405						410	
	aca	gga	ctt	ggg	ctg	ctt	cag	gct	gaa	cgc	tgg	gag	acc	act	tcc	tgt	1462
	Thr	Gly	Leu	Gly	Leu	Leu	Gln	Ala	Glu	Arg	Trp	Glu	Thr	Thr	Ser	Cys	
60					415					420						425	
	gct	gac	aca	cag	agc	tcc	cat	ctc	cat	gaa	gac	cgc	aca	gcg	cgt	gta	1510
	Ala	Asp	Thr	Gln	Ser	Ser	His	Leu	His	Glu	Asp	Arg	Thr	Ala	Arg	Val	
			430					435								440	

```

agc cag ccc agc tgacctaaag cgacatgaga ctactagaaa gaaacgacac      1562
Ser Gln Pro Ser
      445
ccttcccccac gccccacag ctactccaac ccaaacaaca accaagccag tttaatggta      1622
5 ggaatttgta ttttttgcct ttgttcagaa tacatgacat tggtaaataa aaaaaaaaaa      1682
aaaaaa      1688

<210> 43
<211> 1942
10 <212> DNA
    <213> Homo sapiens

<220>
<221> CDS
15 <222> 334..1551

<220>
<221> sig_peptide
<222> 334..426
20 <223> Von Heijne matrix
    score 4.0554926521937
    seq TVFLLVTLQALDT/VE

<400> 43
25 gctcataggg agaaaggaag ctgCGctgCG ttctgCGgga cgaaccctgc tccgCGcgag      60
   aatttttttg attccttctt atttggagaa atctccagct gctctgatca tagcctaaga      120
   agactgcatg ctgcttcctc tcgatgccaa gccagaccct ctcacaacct cggatctcag      180
   tccttcatgg agacctgggc ccagcaggaa tggcagtgca ggaaattggc gccagatgg      240
   ttcttccatg tgaagttgtc tcgggctctg ggctgacgag agaacacctg gtaaccaggt      300
30 tagccctctg tcagtcaccc agggcagggc agc atg gtg cgg att cag agg agg      354
                                   Met Val Arg Ile Gln Arg Arg
                                   -30 -25

   aag ctt ttg gca tct tgc ctg tgc gtc aca gcc acc gtc ttt ctg ctt      402
   Lys Leu Leu Ala Ser Cys Leu Cys Val Thr Ala Thr Val Phe Leu Leu
35   -20 -15 -10
   gtc aca ctc cag gcc ttg gat acc gtt gag aat cta atg aaa gtc acg      450
   Val Thr Leu Gln Ala Leu Asp Thr Val Glu Asn Leu Met Lys Val Thr
                                   -5 1 5
   ggc cct ccc cag gga gtt aca gac tcc atg caa tgc ttc aat gat cag      498
40   Gly Pro Pro Gln Gly Val Thr Asp Ser Met Gln Cys Phe Asn Asp Gln
   10 15 20
   tgg cct tta tct aac acc agg agc agc gag cac ata aaa gag gtc atg      546
   Trp Pro Leu Ser Asn Thr Arg Ser Ser Glu His Ile Lys Glu Val Met
   25 30 35 40
45   gtt gag ctg ggg aag ttt gaa agg aag gag ttt aaa agt tcc agt ttg      594
   Val Glu Leu Gly Lys Phe Glu Arg Lys Glu Phe Lys Ser Ser Ser Leu
   45 50 55
   caa gat gga cat aca aaa atg gag gaa gca cct acg cat ctt aat tca      642
   Gln Asp Gly His Thr Lys Met Glu Glu Ala Pro Thr His Leu Asn Ser
   60 65 70
   ttt ctt aag aaa gaa gga ttg acc ttc aac agg aaa aga aaa tgg gaa      690
   Phe Leu Lys Lys Glu Gly Leu Thr Phe Asn Arg Lys Arg Lys Trp Glu
   75 80 85
   ttg gac agc tac ccc att atg ctc tgg tgg tcc ccg ctg acg ggg gag      738
55   Leu Asp Ser Tyr Pro Ile Met Leu Trp Trp Ser Pro Leu Thr Gly Glu
   90 95 100
   act ggg agg tta ggc caa tgt gga gca gat gct tgt ttc ttc acc atc      786
   Thr Gly Arg Leu Gly Gln Cys Gly Ala Asp Ala Cys Phe Phe Thr Ile
   105 110 115 120
60   aac cgg acc tac ctc cat cat cac atg acc aaa gca ttc ctc ttc tat      834
   Asn Arg Thr Tyr Leu His His His Met Thr Lys Ala Phe Leu Phe Tyr
   125 130 135
   ggt act gac ttt aac ata gat agc tta cct ctg cct cgg aaa gcc cat      882

```

	Gly	Thr	Asp	Phe	Asn	Ile	Asp	Ser	Leu	Pro	Leu	Pro	Arg	Lys	Ala	His	
				140					145					150			
	cat	gac	tgg	gct	gtt	ttt	cat	gaa	gag	tcc	ccg	aaa	aac	aat	tat	aag	930
	His	Asp	Trp	Ala	Val	Phe	His	Glu	Glu	Ser	Pro	Lys	Asn	Asn	Tyr	Lys	
5			155					160					165				
	ctc	ttt	cat	aaa	cca	gtg	atc	acc	ttg	ttc	aac	tac	act	gcc	acg	ttc	978
	Leu	Phe	His	Lys	Pro	Val	Ile	Thr	Leu	Phe	Asn	Tyr	Thr	Ala	Thr	Phe	
			170					175					180				
	agc	agg	cat	tcc	cac	ttg	cca	cta	act	acc	caa	tac	ttg	gag	agc	att	1026
10	Ser	Arg	His	Ser	His	Leu	Pro	Leu	Thr	Thr	Gln	Tyr	Leu	Glu	Ser	Ile	
	185					190					195					200	
	gaa	gtc	ctg	aag	tca	ctc	cga	tac	cta	gtt	cct	ttg	cag	tcc	aaa	aac	1074
	Glu	Val	Leu	Lys	Ser	Leu	Arg	Tyr	Leu	Val	Pro	Leu	Gln	Ser	Lys	Asn	
					205					210					215		
15	aag	ctt	aga	aaa	aga	ctt	gct	ccg	ctg	gtg	tat	gta	cag	tca	tac	tgt	1122
	Lys	Leu	Arg	Lys	Arg	Leu	Ala	Pro	Leu	Val	Tyr	Val	Gln	Ser	Tyr	Cys	
					220				225					230			
	gac	cca	cca	tca	gac	agg	gac	agc	tat	gtt	cgc	gag	ctg	atg	act	tac	1170
	Asp	Pro	Pro	Ser	Asp	Arg	Asp	Ser	Tyr	Val	Arg	Glu	Leu	Met	Thr	Tyr	
20			235					240					245				
	atc	gag	gtc	gat	tcc	tat	ggg	gaa	tgt	tta	cga	aac	aaa	gac	ctc	cct	1218
	Ile	Glu	Val	Asp	Ser	Tyr	Gly	Glu	Cys	Leu	Arg	Asn	Lys	Asp	Leu	Pro	
			250				255					260					
	cag	cag	ctg	aaa	aat	cca	gcc	tct	atg	gat	gcc	gat	ggc	ttt	tat	agg	1266
25	Gln	Gln	Leu	Lys	Asn	Pro	Ala	Ser	Met	Asp	Ala	Asp	Gly	Phe	Tyr	Arg	
	265					270				275						280	
	atc	att	gca	cag	tat	aag	ttt	atc	cta	gct	ttt	gag	aat	gca	gtt	tgt	1314
	Ile	Ile	Ala	Gln	Tyr	Lys	Phe	Ile	Leu	Ala	Phe	Glu	Asn	Ala	Val	Cys	
					285				290					295			
30	gat	gac	tac	atc	act	gag	aag	ttc	tgg	agg	cca	ctg	aaa	ctg	ggg	gta	1362
	Asp	Asp	Tyr	Ile	Thr	Glu	Lys	Phe	Trp	Arg	Pro	Leu	Lys	Leu	Gly	Val	
					300				305					310			
	gtc	cct	gta	tat	tac	gga	tcc	ccc	agc	atc	aca	gac	tgg	ctt	cca	agt	1410
	Val	Pro	Val	Tyr	Tyr	Gly	Ser	Pro	Ser	Ile	Thr	Asp	Trp	Leu	Pro	Ser	
35			315					320					325				
	aac	aaa	agt	gct	att	ctt	gta	tca	gaa	ttt	tct	cac	ccc	agg	gaa	ctg	1458
	Asn	Lys	Ser	Ala	Ile	Leu	Val	Ser	Glu	Phe	Ser	His	Pro	Arg	Glu	Leu	
			330				335					340					
	gca	agt	tac	atc	aga	cga	ctg	gat	tct	gat	gac	aga	ttg	tat	gag	gcc	1506
40	Ala	Ser	Tyr	Ile	Arg	Arg	Leu	Asp	Ser	Asp	Asp	Arg	Leu	Tyr	Glu	Ala	
	345					350				355						360	
	tat	gta	gaa	tgg	aag	ctg	aag	ggg	aga	tct	cta	acc	agc	gac	ttc		1551
	Tyr	Val	Glu	Trp	Lys	Leu	Lys	Gly	Arg	Ser	Leu	Thr	Ser	Asp	Phe		
					365				370					375			
45	tgacagctct	caggggaacgg	aaatgggggag	tgcaagacgt	caaccaggac	aattacatcg											1611
	atgcatttga	gtgtatgggtg	tgcaccaagg	tgtgggctaa	tatcaggctt	caggaaaagg											1671
	gcttaccacc	caaaagatgg	gaggcagaag	ataccacact	gagttgcccc	gagccacag											1731
	tgtttgcttt	ctcaccactc	cggactccac	ctttgagctc	tttgcgagag	atgtggattt											1791
	ccagctttga	acaatccaag	aaagaagccc	aggcactaag	gtggctgggt	gataggaatc											1851
50	aaaacttttc	atctcaagag	ttttggggcc	tagtattcaa	ggactgattt	caaaaatgat											1911
	cagaatgaaa	cagaaaaaaa	aaaaaaaaaa	a													1942

<210> 44
 <211> 1657
 55 <212> DNA
 <213> Homo sapiens

 <220>
 <221> CDS
 60 <222> 72..986

<220>
 <221> sig_peptide

<222> 72..149

<223> Von Heijne matrix

score 6.33091407142367

seq GVGLVTLGLAVG/SY

5

<400> 44

	ctccgacccg	ccccgcggcg	cattgtggga	tctgtcggct	tgtcaggtgg	tggaggaaaa	60
	ggcgctccgt	c	atg ggg atc	cag acg agc	ccc gtc ctg	ctg gcc tcc	ctg 110
			Met Gly Ile	Gln Thr Ser	Pro Val Leu	Leu Ala Ser	Leu
10			-25		-20	-15	
	ggg gtc ggg	ctg gtc act	ctg ctc ggc	ctg gct gtc	ggc tcc tac	ttg	158
	Gly Val Gly	Leu Val Thr	Leu Leu Gly	Leu Ala Val	Gly Ser Tyr	Leu	
		-10		-5		1	
	gtt cgg agg	tcc cgc cgg	cct cag gtc	act ctc ctg	gac ccc aat	gaa	206
15	Val Arg Arg	Ser Arg Arg	Pro Gln Val	Thr Leu Leu	Asp Pro Asn	Glu	
	5		10		15		
	aag tac ctg	cta cga ctg	cta gac aag	acg act gtg	agc cac aac	acc	254
	Lys Tyr Leu	Leu Arg Leu	Leu Asp Lys	Thr Thr Val	Ser His Asn	Thr	
	20		25		30	35	
20	aag agg ttc	cgc ttt gcc	ctg ccc acc	gcc cac cac	act ctg ggg	ctg	302
	Lys Arg Phe	Arg Phe Ala	Leu Pro Thr	Ala His His	Thr Leu Gly	Leu	
		40		45		50	
	cct gtg ggc	aaa cat atc	tac ctc tcc	acc cga att	gat ggc agc	ctg	350
	Pro Val Gly	Lys His Ile	Tyr Leu Ser	Thr Arg Ile	Asp Gly Ser	Leu	
25		55		60		65	
	gtc atc agg	cca tac act	cct gtc acc	agt gat gag	gat caa ggc	tat	398
	Val Ile Arg	Pro Tyr Thr	Pro Val Thr	Ser Asp Glu	Asp Gln Gly	Tyr	
	70		75		80		
	gtg gat ctt	gtc atc aag	gtc tac ctg	aag ggt gtg	cac ccc aaa	ttt	446
30	Val Asp Leu	Val Ile Lys	Val Tyr Leu	Lys Gly Val	His Pro Lys	Phe	
	85		90		95		
	cct gag gga	ggg aag atg	tct cag tac	ctg gat agc	ctg aag gtt	ggg	494
	Pro Glu Gly	Gly Lys Met	Ser Gln Tyr	Leu Asp Ser	Leu Lys Val	Gly	
	100		105		110	115	
35	gat gtg gtg	gag ttt cgg	ggg cca agc	ggg ttg ctc	act tac act	gga	542
	Asp Val Val	Glu Phe Arg	Gly Pro Ser	Gly Leu Leu	Thr Tyr Thr	Gly	
		120		125		130	
	aaa ggg cat	ttt aac att	cag ccc aac	aag aaa tct	cca cca gaa	ccc	590
	Lys Gly His	Phe Asn Ile	Gln Pro Asn	Lys Lys Ser	Pro Pro Glu	Pro	
40		135		140		145	
	cga gtg gcg	aag aaa ctg	gga atg att	gcc gcc ggg	aca gga atc	acc	638
	Arg Val Ala	Lys Lys Leu	Gly Met Ile	Ala Gly Gly	Thr Gly Ile	Thr	
	150		155		160		
	cca atg cta	cag ctg atc	cgg gcc atc	ctg aaa gtc	cct gaa gat	cca	686
45	Pro Met Leu	Gln Leu Ile	Arg Ala Ile	Leu Lys Val	Pro Glu Asp	Pro	
	165		170		175		
	acc cag tgc	ttt ctg ctt	ttt gcc aac	cag aca gaa	aag gat atc	atc	734
	Thr Gln Cys	Phe Leu Leu	Phe Ala Asn	Gln Thr Glu	Lys Asp Ile	Ile	
	180		185		190	195	
50	ttg cgg gag	gac tta gag	gaa ctg cag	gcc cgc tat	ccc aat cgc	ttt	782
	Leu Arg Glu	Asp Leu Glu	Glu Leu Gln	Ala Arg Tyr	Pro Asn Arg	Phe	
		200		205		210	
	aag ctc tgg	ttc act ctg	gat cat ccc	cca aaa gat	tgg gcc tac	agc	830
	Lys Leu Trp	Phe Thr Leu	Asp His Pro	Pro Lys Asp	Trp Ala Tyr	Ser	
55		215		220		225	
	aag ggc ttt	gtg act gcc	gac atg atc	cgg gaa cac	ctg ccc gct	cca	878
	Lys Gly Phe	Val Thr Ala	Asp Met Ile	Arg Glu His	Leu Pro Ala	Pro	
	230		235		240		
	ggg gat gat	gtg ctg gta	ctg ctt tgt	ggg cca ccc	cca atg gtg	cag	926
60	Gly Asp Asp	Val Leu Val	Leu Leu Cys	Gly Pro Pro	Pro Pro Met	Val Gln	
	245		250		255		
	ctg gcc tgc	cat ccc aac	ttg gac aaa	ctg ggc tac	tca caa aag	atg	974
	Leu Ala Cys	His Pro Asn	Leu Asp Lys	Leu Gly Tyr	Ser Gln Lys	Met	

260 265 270 275 1026
cga ttc acc tac tgagcatcct ccagcttccc tgggtgctggt cgctgcagtt
Arg Phe Thr Tyr
gttccccatc agtactcaag cactataagc cttagattcc tttcctcaga gtttcaggtt 1086
5 ttttcagtta catctagagc tgaaatctgg atagtacctg caggaacaat attcctgtag 1146
ccatggaaga gggccaaggc tcagtcactc cttggatggc ctcttaaate tccccgtggc 1206
aacaggtcca ggagaggccc atggagcagt ctcttccatg gagtaagaag gaagggagca 1266
tgtacgcttg gtccaagatt ggctagttcc ttgatagcat cttactctca ccttctttgt 1326
gtctgtgatg aaaggaacag tctgtgcaat gggttttact taaacttcac tgttcaacct 1386
10 atgagcaaat ctgtatgtgt gagtataagt tgagcatagc atacttccag aggtggtctt 1446
atggagatgg caagaaagga ggaaatgatt tcttcagatc tcaaaggagt ctgaaatata 1506
atatttctgt gtgtgtctct ctacagcccct gccaggcta gagggaaaca gctactgata 1566
atcgaaaact gctgtttgtg gcaggaaccc ctggctgtgc aaataaatgg ggctgaggcc 1626
cctgtgtgat attaaaaaaa aaaaaaaaaa a 1657

15
<210> 45
<211> 1733
<212> DNA
<213> Homo sapiens

20
<220>
<221> CDS
<222> 157..1482

25
<220>
<221> sig_peptide
<222> 157..219
<223> Von Heijne matrix
score 11.6921972463885

30
seq LLLCLALSGAAET/KP

<400> 45
aaagaaaagt cggcagcaga gggaacaggg aagaaaccta aaggctgcag gctgccaggt 60
gtgcttggag agcccccttc ttccgccggg cctcgcaagc agcgtaggac tgtggagaag 120
35 ggcggtgggc aaggagggaa ctcgagagca gcctcc atg ggc aca cag gag ggc 174
Met Gly Thr Gln Glu Gly
-20
tgg tgc ctg ctg ctc tgc ctg gct cta tct gga gca gca gaa acc aag 222
Trp Cys Leu Leu Leu Cys Leu Ala Leu Ser Gly Ala Ala Glu Thr Lys
40 -15 -10 -5 1
ccc cac cca gca gag ggg cag ttg cgg gca gtg gac gtg gtc cta gac 270
Pro His Pro Ala Glu Gly Gln Leu Arg Ala Val Asp Val Val Leu Asp
5 10 15
tgc ttc ctg gcg aag gac ggt gcg cac cgt gga gct ctc gcc agc agt 318
45 Cys Phe Leu Ala Lys Asp Gly Ala His Arg Gly Ala Leu Ala Ser Ser
20 25 30
gag gac agg gca agg gcc tcc ctt gtg ctg aag cag gtg cca gtg ctg 366
Glu Asp Arg Ala Arg Ala Ser Leu Val Leu Lys Gln Val Pro Val Leu
35 40 45
50 gac gat ggc tcc ctg gag gac ttc acc gat ttc caa ggg ggc aca ctg 414
Asp Asp Gly Ser Leu Glu Asp Phe Thr Asp Phe Gln Gly Gly Thr Leu
50 55 60 65
gcc caa gat gac cca cct att atc ttt gag gcc tca gtg gac ctg gtc 462
Ala Gln Asp Asp Pro Pro Ile Ile Phe Glu Ala Ser Val Asp Leu Val
55 70 75 80
cag att ccc cag gcc gag gcc ttg ctc cat gct gac tgc agt ggg aag 510
Gln Ile Pro Gln Ala Glu Ala Leu Leu His Ala Asp Cys Ser Gly Lys
85 90 95
60 gag gtg acc tgt gag atc tcc cgc tac ttt ctc cag atg aca gag acc 558
Glu Val Thr Cys Glu Ile Ser Arg Tyr Phe Leu Gln Met Thr Glu Thr
100 105 110
act gtt aag aca gca gct tgg ttc atg gcc aac atg cag gtc tct gga 606
Thr Val Lys Thr Ala Ala Trp Phe Met Ala Asn Met Gln Val Ser Gly

		115				120				125								
		ggg gga cst agc atc tcc ttg gtg atg aag act ccc agg gtc acc aag																654
		Gly Gly Xaa Ser Ile Ser Leu Val Met Lys Thr Pro Arg Val Thr Lys																
		130				135				140							145	
5		aat gag gcg ctc tgg cac ccg acg ctg aac ttg cca ctg agc ccc cag																702
		Asn Glu Ala Leu Trp His Pro Thr Leu Asn Leu Pro Leu Ser Pro Gln																
				150					155								160	
		ggg act gtg cga act gca gtg gag ttc cag gtg atg aca cag acc caa																750
		Gly Thr Val Arg Thr Ala Val Glu Phe Gln Val Met Thr Gln Thr Gln																
10				165					170								175	
		tcc ctg agc ttc ctg ctg ggg tcc tca gcc tcc ttg gac tgt ggc ttc																798
		Ser Leu Ser Phe Leu Leu Gly Ser Ser Ala Ser Leu Asp Cys Gly Phe																
				180					185								190	
		tcc atg gca ccg ggc ttg gac ctc atc agt gtg gag tgg cga ctg cag																846
15		Ser Met Ala Pro Gly Leu Asp Leu Ile Ser Val Glu Trp Arg Leu Gln																
				195					200								205	
		cac aag ggc agg ggt cag ttg gtg tac agc tgg acc gca ggg cag ggg																894
		His Lys Gly Arg Gly Gln Leu Val Tyr Ser Trp Thr Ala Gly Gln Gly																
				210					215								220	
20		cag gct gtg cgg aag ggc gct acc ctg gag cct gca caa ctg ggc atg																942
		Gln Ala Val Arg Lys Gly Ala Thr Leu Glu Pro Ala Gln Leu Gly Met																
				230					235								240	
		gcc agg gat gcc tcc ctc acc ctg ccc ggc ctc act ata cag gac gag																990
		Ala Arg Asp Ala Ser Leu Thr Leu Pro Gly Leu Thr Ile Gln Asp Glu																
25				245					250								255	
		ggg acc tac att tgc cag atc acc acc tct ctg tac cga gct cag cag																1038
		Gly Thr Tyr Ile Cys Gln Ile Thr Ser Leu Tyr Arg Ala Gln Gln																
				260					265								270	
		atc atc cag ctc aac atc caa gct tcc cct aaa gta cga ctg agc ttg																1086
30		Ile Ile Gln Leu Asn Ile Gln Ala Ser Pro Lys Val Arg Leu Ser Leu																
				275					280								285	
		gca aac gaa gct ctg ctg ccc acc ctc atc tgc gac att gct ggc tat																1134
		Ala Asn Glu Ala Leu Leu Pro Thr Leu Ile Cys Asp Ile Ala Gly Tyr																
				290					295								300	
35		tac cct ctg gat gtg gtg gtg acg tgg acc cga gag gag ctg ggt gga																1182
		Tyr Pro Leu Asp Val Val Val Thr Trp Thr Arg Glu Glu Leu Gly Gly																
				310					315								320	
		tcc cca gcc caa gtc tct ggt gcc tcc ttc tcc agc ctc agg caa agc																1230
40		Ser Pro Ala Gln Val Ser Gly Ala Ser Phe Ser Ser Ser Leu Arg Gln Ser																
				325					330								335	
		gtg gca ggc acc tac agc atc tcc tcc tct ctc acc gca gaa cct ggc																1278
		Val Ala Gly Thr Tyr Ser Ile Ser Ser Ser Leu Thr Ala Glu Pro Gly																
				340					345								350	
		tct gca ggt gcc act tac acc tgc cag gtc aca cac atc tct ctg gag																1326
45		Ser Ala Gly Ala Thr Tyr Thr Cys Gln Val Thr His Ile Ser Leu Glu																
				355					360								365	
		gag ccc ctt ggg gcc agc acc cag gtt gtc cca cca gag cgg aga aca																1374
		Glu Pro Leu Gly Ala Ser Thr Gln Val Val Pro Pro Glu Arg Arg Thr																
				370					375								380	
50		gcc ttg gga gtc atc ttt gcc agc agt ctc ttc ctt ctt gca ctg atg																1422
		Ala Leu Gly Val Ile Phe Ala Ser Ser Leu Phe Leu Leu Ala Leu Met																
				390					395								400	
		ttc ctg ggg ctt cag aga cgg caa gca cct aca gga ctt ggg ctg ctt																1470
		Phe Leu Gly Leu Gln Arg Arg Gln Ala Pro Thr Gly Leu Gly Leu Leu																
55				405					410								415	
		cag gct gaa cgc taggagacca cttcctgtgc tgacacacag agtcccatc																1522
		Gln Ala Glu Arg																
				420														
60		tccatgaaga ccgcacagcg cgtgtaagcc agcccagctg acctaaagcg acatgagact																1582
		actagaaga aacgacaccc ttccccaagc ccccacagct actccaaccc aaacaacaac																1642
		caagccagtt taatggtagg aatttgtatt ttttgccctt gttcagaata catgacattg																1702
		gtaaatatgc cacaaaaaaaa aaaaaaaaaa a																1733


```

<210> 46
<211> 1871
<212> DNA
<213> Homo sapiens
5
<220>
<221> CDS
<222> 195..1052

10 <220>
    <221> sig_peptide
    <222> 195..338
    <223> Von Heijne matrix
        score 3.50178852790004
15     seq LGV FVVCHQLSSS/LN

<400> 46
agtgactgcc gggagtcctg caggggcggg gcggcgccaa gcgcagggag cccggctgag      60
tggcagccca gattgaagat ggatacgtga caatcccagg gaccgctgca ctgacttcat      120
20 ttccttagac aagacacagt gtagggcccg gcccggtgtg gccccaggac tcctttggaa      180
tatagctgtg gaca atg aat cct gcg agc gat ggg ggc aca tca gag agc      230
                Met Asn Pro Ala Ser Asp Gly Gly Thr Ser Glu Ser
                -45                                -40

25 att ttt gac ctg gac tat gca tcc tgg ggg atc cgc tcc acg ctg atg      278
Ile Phe Asp Leu Asp Tyr Ala Ser Trp Gly Ile Arg Ser Thr Leu Met
    -35                                -30                                -25

    gtc gct ggc ttt gtc ttc tac ttg ggc gtc ttt gtg gtc tgc cac cag      326
    Val Ala Gly Phe Val Phe Tyr Leu Gly Val Phe Val Val Cys His Gln
    -20                                -15                                -10                                -5

30 ctg tcc tct tcc ctg aat gcc act tac cgt tct ttg gtg gcc aga gag      374
Leu Ser Ser Ser Leu Asn Ala Thr Tyr Arg Ser Leu Val Ala Arg Glu
                1                5                10

    aag gtc ttc tgg gac ctg gcg gcc acg cgt gca gtc ttt ggt gtt cag      422
    Lys Val Phe Trp Asp Leu Ala Ala Thr Arg Ala Val Phe Gly Val Gln
    15                20                25

35 agc aca gcc gca ggc ctg tgg gct ctg ctg ggg gac cct gtg ctg cat      470
Ser Thr Ala Ala Gly Leu Trp Ala Leu Leu Gly Asp Pro Val Leu His
    30                35                40

    gcc gac aag gcg cgt ggc cag cag aac tgg tgc tgg ttt cac atc acg      518
    Ala Asp Lys Ala Arg Gly Gln Gln Asn Trp Cys Trp Phe His Ile Thr
    45                50                55                60

    aca gca acg gga ttc ttt tgc ttt gaa aat gtt gca gtc cac ctg tcc      566
    Thr Ala Thr Gly Phe Phe Cys Phe Glu Asn Val Ala Val His Leu Ser
    65                70                75

45 aac ttg atc ttc cgg aca ttt gac ttg ttt ctg gtt atc cac cat ctc      614
Asn Leu Ile Phe Arg Thr Phe Asp Leu Phe Leu Val Ile His His Leu
    80                85                90

    ttt gcc ttt ctt ggg ttt ctt ggc tgc ttg gtc aat ctc caa gct ggc      662
    Phe Ala Phe Leu Gly Phe Leu Gly Cys Leu Val Asn Leu Gln Ala Gly
    95                100                105

50 cac tat cta gct atg acc acg ttg ctc ctg gag atg agc acg ccc ttt      710
His Tyr Leu Ala Met Thr Thr Leu Leu Leu Glu Met Ser Thr Pro Phe
    110                115                120

    acc tgc gtt tcc tgg atg ctc tta aag gcg ggc tgg tcc gag tct ctg      758
    Thr Cys Val Ser Trp Met Leu Leu Lys Ala Gly Trp Ser Glu Ser Leu
    125                130                135                140

    ttt tgg aag ctc aac cag tgg ctg atg att cac atg ttt cac tgc cgc      806
    Phe Trp Lys Leu Asn Gln Trp Leu Met Ile His Met Phe His Cys Arg
    145                150                155

60 atg gtt cta acc tac cac atg tgg tgg gtg tgt ttc tgg cac tgg gac      854
Met Val Leu Thr Tyr His Met Trp Trp Val Cys Phe Trp His Trp Asp
    160                165                170

    ggc ctg gtc agc agc ctg tat ctg cct cat ttg aca ctg ttc ctt gtc      902

```

Gly Leu Val Ser Ser Leu Tyr Leu Pro His Leu Thr Leu Phe Leu Val
 175 180 185
 gga ctg gct ctg ctt acg cta atc att aat cca tat tgg acc cat aag 950
 Gly Leu Ala Leu Leu Thr Leu Ile Ile Asn Pro Tyr Trp Thr His Lys
 5 190 195 200
 aag act cag cag ctt ctc aat ccg gtg gac tgg aac ttc gca cag cca 998
 Lys Thr Gln Gln Leu Leu Asn Pro Val Asp Trp Asn Phe Ala Gln Pro
 205 210 215 220
 gaa gcc aag agc agg cca gaa ggc aac ggg cag ctg ctg cgg aag aag 1046
 10 Glu Ala Lys Ser Arg Pro Glu Gly Asn Gly Gln Leu Leu Arg Lys Lys
 225 230 235
 agg cca tagctgctcc agccggggct ccggggcggc agcagagctg gcacaccgat 1102
 Arg Pro
 tctgggaagc cccgcgaatg atggcttttg aattaatgag gcagtgaatg ttttgtgttt 1162
 15 acttctaagg gaaatactaa ctttctttcg cattagtatt aattttgaag tagctacaaa 1222
 gtatttttaa gaaattataa ttttatgact gtctggcagg ctctgtcagt ttagccgcgc 1282
 cggaccgtgt caagcatcta ggagaggagt ccatggtgtc caggcatcgg ggcgtcacac 1342
 ctggtgagga gtgggtggc tttgaatgct ggaaatggct tcatagtga gtgcctccca 1402
 cagggcggtt gggtcagcgt tgactctttc cagctgcaca ctcatatgcc gtgtgtctta 1462
 20 ttcagaagtc acattctttt cagttggaga gaattgggct aagatagaaa ataacatgat 1522
 ttgttcctta ttaaagtttc ccagcgtatg aaattctaag ctgggtgggg tggctcacac 1582
 ccgacgtaat cccagcacgt tgggaggccg aggcagggtg atcacttgag gccaggagtt 1642
 cgagaccagc ctggtcaaga tggtgaaacc ccatctctac taaaattaca aaaattagcc 1702
 ggggtgtcgtg gcacacacct gtaatcccag ctatttggga ggccaaggca ggagaattgc 1762
 25 ctgaacccgg gaggcggagg ttgcagttag ctgagatcgc accactgcac tccagcactc 1822
 cagcctgggt gacggagcaa cactctctcg caaaaaaaaa aaaaaaaaaa 1871

 <210> 47
 <211> 1523
 30 <212> DNA
 <213> Homo sapiens

 <220>
 <221> CDS
 35 <222> 217..1410

 <220>
 <221> sig_peptide
 <222> 217..279
 40 <223> Von Heijne matrix
 score 5.8172934575094
 seq ALLWAQEVGQVLA/GR

 <400> 47
 45 acttccccgg gagccggaag tcccgctctca cggttgccct ggcagcgcgc gaggctggtg 60
 agtcggcagc cctgtggcag ccggcgggct ggtttccatg gttgcacgat taggaaccac 120
 cagctgctgc atcccattggc caggggtggc gtccagggtg cagagcagct aggaacgcaa 180
 ggcctgaacc tggggccaga caccctctc ccggcc atg gtc aac gac cct cca 234
 Met Val Asn Asp Pro Pro
 50 -20
 gta cct gcc tta ctg tgg gcc cag gag gtg ggc caa gtc ttg gca ggc 282
 Val Pro Ala Leu Leu Trp Ala Gln Glu Val Gly Gln Val Leu Ala Gly
 -15 -10 -5 1
 cgt gcc cgc agg ctg ctg ctg cag ttt ggg gtg ctc ttc tgc acc atc 330
 55 Arg Ala Arg Arg Leu Leu Leu Gln Phe Gly Val Leu Phe Cys Thr Ile
 5 10 15
 ctc ctt ttg ctc tgg gtg tct gtc ttc ctc tat ggc tcc ttc tac tat 378
 Leu Leu Leu Leu Trp Val Ser Val Phe Leu Tyr Gly Ser Phe Tyr Tyr
 20 25 30
 60 tcc tat atg ccg aca gtc agc cac ctc agc cct gtg cat ttc tac tac 426
 Ser Tyr Met Pro Thr Val Ser His Leu Ser Pro Val His Phe Tyr Tyr
 35 40 45
 agg acc gac tgt gat tcc tcc acc acc tca ctc tgc tcc ttc cct gtt 474

	Arg	Thr	Asp	Cys	Asp	Ser	Ser	Thr	Thr	Ser	Leu	Cys	Ser	Phe	Pro	Val	
	50					55					60					65	
	gcc	aat	gtc	tcg	ctg	act	aag	ggg	gga	cgt	gat	cgg	gtg	ctg	atg	tat	522
	Ala	Asn	Val	Ser	Leu	Thr	Lys	Gly	Gly	Arg	Asp	Arg	Val	Leu	Met	Tyr	
5					70					75					80		
	gga	cag	ccg	tat	cgt	gtt	acc	tta	gag	ctt	gag	ctg	cca	gag	tcc	cct	570
	Gly	Gln	Pro	Tyr	Arg	Val	Thr	Leu	Glu	Leu	Glu	Leu	Pro	Glu	Ser	Pro	
				85					90				95				
	gtg	aat	caa	gat	ttg	ggc	atg	ttc	ttg	gtc	acc	att	tcc	tgc	tac	acc	618
10	Val	Asn	Gln	Asp	Leu	Gly	Met	Phe	Leu	Val	Thr	Ile	Ser	Cys	Tyr	Thr	
			100					105					110				
	aga	ggg	ggc	cga	atc	atc	tcc	act	tct	tcg	cgt	tcg	gtg	atg	ctg	cat	666
	Arg	Gly	Gly	Arg	Ile	Ile	Ser	Thr	Ser	Ser	Arg	Ser	Val	Met	Leu	His	
		115					120					125					
15	tac	cgc	tca	gac	ctg	ctc	cag	atg	ctg	gac	aca	ctg	gtc	ttc	tct	agc	714
	Tyr	Arg	Ser	Asp	Leu	Leu	Gln	Met	Leu	Asp	Thr	Leu	Val	Phe	Ser	Ser	
					130		135				140					145	
	ctc	ctg	cta	ttt	ggc	ttt	gca	gag	cag	aag	cag	ctg	ctg	gag	gtg	gaa	762
	Leu	Leu	Leu	Phe	Gly	Phe	Ala	Glu	Gln	Lys	Gln	Leu	Leu	Glu	Val	Glu	
20					150					155					160		
	ctc	tac	gca	gac	tat	aga	gag	aac	tcg	tac	gtg	ccg	acc	act	gga	gcg	810
	Leu	Tyr	Ala	Asp	Tyr	Arg	Glu	Asn	Ser	Tyr	Val	Pro	Thr	Thr	Gly	Ala	
				165					170					175			
25	atc	att	gag	atc	cac	agc	aag	cgc	atc	cag	ctg	tat	gga	gcc	tac	ctc	858
	Ile	Ile	Glu	Ile	His	Ser	Lys	Arg	Ile	Gln	Leu	Tyr	Gly	Ala	Tyr	Leu	
			180					185					190				
	cgc	atc	cac	gcg	cac	ttc	act	ggg	ctc	aga	tac	ctg	cta	tac	aac	ttc	906
	Arg	Ile	His	Ala	His	Phe	Thr	Gly	Leu	Arg	Tyr	Leu	Leu	Tyr	Asn	Phe	
			195				200					205					
30	ccg	atg	acc	tgc	gcc	ttc	ata	ggg	gtt	gcc	agc	aac	ttc	acc	ttc	ctc	954
	Pro	Met	Thr	Cys	Ala	Phe	Ile	Gly	Val	Ala	Ser	Asn	Phe	Thr	Phe	Leu	
					210		215				220					225	
	agc	gtc	atc	gtg	ctc	ttc	agc	tac	atg	cag	tgg	gtg	tgg	ggg	ggc	atc	1002
	Ser	Val	Ile	Val	Leu	Phe	Ser	Tyr	Met	Gln	Trp	Val	Trp	Gly	Gly	Ile	
35					230				235						240		
	tgg	ccc	cga	cac	cgc	ttc	tct	ttg	cag	gtt	aac	atc	cga	aaa	aga	gac	1050
	Trp	Pro	Arg	His	Arg	Phe	Ser	Leu	Gln	Val	Asn	Ile	Arg	Lys	Arg	Asp	
				245					250					255			
40	aat	tcc	cgg	aag	gaa	gtc	caa	cga	agg	atc	tct	gct	cat	cag	cca	ggg	1098
	Asn	Ser	Arg	Lys	Glu	Val	Gln	Arg	Arg	Ile	Ser	Ala	His	Gln	Pro	Gly	
				260				265					270				
	cct	gaa	ggc	cag	gag	gag	tca	act	ccg	caa	tca	gat	gtt	aca	gag	gat	1146
	Pro	Glu	Gly	Gln	Glu	Glu	Ser	Thr	Pro	Gln	Ser	Asp	Val	Thr	Glu	Asp	
			275				280					285					
45	ggg	gag	agc	cct	gaa	gat	ccc	tca	ggg	aca	gag	ggg	cag	ctg	tcc	gag	1194
	Gly	Glu	Ser	Pro	Glu	Asp	Pro	Ser	Gly	Thr	Glu	Gly	Gln	Leu	Ser	Glu	
					290		295				300					305	
	gag	gag	aaa	cca	gat	cag	cag	ccc	ctg	agc	gga	gaa	gag	gag	cta	gag	1242
	Glu	Glu	Lys	Pro	Asp	Gln	Gln	Pro	Leu	Ser	Gly	Glu	Glu	Glu	Leu	Glu	
50					310				315						320		
	cct	gag	gcc	agt	gat	ggg	tca	ggc	tcc	tgg	gaa	gat	gca	gct	ttg	ctg	1290
	Pro	Glu	Ala	Ser	Asp	Gly	Ser	Gly	Ser	Trp	Glu	Asp	Ala	Ala	Leu	Leu	
				325					330					335			
55	acg	gag	gcc	aac	ctg	cct	gct	cct	cct	gct	tct	gct	tct	gcc	cct		1338
	Thr	Glu	Ala	Asn	Leu	Pro	Ala	Pro	Ala	Pro	Ala	Ser	Ala	Ser	Ala	Pro	
			340				345						350				
	gtc	cta	gag	act	ctg	ggc	agc	tct	gaa	cct	gct	ggg	ggg	gct	ctc	cga	1386
	Val	Leu	Glu	Thr	Leu	Gly	Ser	Ser	Glu	Pro	Ala	Gly	Gly	Ala	Leu	Arg	
			355				360					365					
60	cag	cgc	ccc	acc	tgc	tct	agt	tcc	tgaagaaaag	gggcagactc	ctcacattcc						1440
	Gln	Arg	Pro	Thr	Cys	Ser	Ser	Ser									
					370		375										
	agcactttcc	cacctgactc	ctctccctc	gtttttcctt	caataaacta	ttttgtgtca											1500

gcttcgaaaa aaaaaaaaaa aaa 1523

<210> 48
 <211> 832
 5 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 10 <222> 103..492

<220>
 <221> sig_peptide
 <222> 103..162
 15 <223> Von Heijne matrix
 score 8.21832452871462
 seq LFFCYLLFTCSG/VE

<400> 48
 20 gtttactcgc tgctgtgccc atctatcagc aggtccggg ctgaagattg cttctcttct 60
 ctccctccaag gtctagtac ggagcccgcg cgcggcgcca cc atg cgg cag aag 114
 Met Arg Gln Lys
 -20

gcg gta tcg ctt ttc ttc tgc tac ctg ctg ctc ttc act tgc agt ggg 162
 25 Ala Val Ser Leu Phe Phe Cys Tyr Leu Leu Leu Phe Thr Cys Ser Gly
 -15 -10 -5
 gtg gag gca ggt aag aaa aag tgc tcg gag agc tcg gac agc ggc tcc 210
 Val Glu Ala Gly Lys Lys Lys Cys Ser Glu Ser Ser Asp Ser Gly Ser
 1 5 10 15

30 ggg ttc tgg aag gcc ctg acc ttc atg gcc gtc gga gga gga ctc gca 258
 Gly Phe Trp Lys Ala Leu Thr Phe Met Ala Val Gly Gly Gly Leu Ala
 20 25 30

gtc gcc ggg ctg ccc gcg ctg ggc ttc acc ggc gcc ggc atc gcg gcc 306
 Val Ala Gly Leu Pro Ala Leu Gly Phe Thr Gly Ala Gly Ile Ala Ala
 35 35 40 45

35 aac tcg gtg gct gcc tcg ctg atg agc tgg tct gcg atc ctg aat ggg 354
 Asn Ser Val Ala Ala Ser Leu Met Ser Trp Ser Ala Ile Leu Asn Gly
 50 55 60

ggc ggc gtg ccc gcc ggg ggg cta gtg gcc acg ctg cag agc ctc ggg 402
 40 Gly Gly Val Pro Ala Gly Gly Leu Val Ala Thr Leu Gln Ser Leu Gly
 65 70 75 80

gct ggt ggc agc agc gtc gtc ata ggt aat att ggt gcc ctg atg ggc 450
 Ala Gly Gly Ser Ser Val Val Ile Gly Asn Ile Gly Ala Leu Met Gly
 85 90 95

45 tac gcc acc cac aag tat ctc gat agt gag gag gat gag gag 492
 Tyr Ala Thr His Lys Tyr Leu Asp Ser Glu Glu Asp Glu Glu
 100 105 110

tagccagcag ctcccagaac ctcttcttcc ttcttggcct aactcttcca gttaggatct 552
 agaactttgc cttttttttt tttttttttt tttttttgag atgggttctc actatattgt 612
 50 ccaggctaga gtgcagtggc tattcacaga tgcgaacata gtacactgca gcctccaact 672
 cctagcctca agtgatcctc ctgtctcaac ctcccaagta ggattacaag catgcgccga 732
 cgatgccccag aatccagaac tttgtctatc actctcccca acaacctaga tgtgaaaaca 792
 gaataaactt caccagaaaa gcaaaaaaaaa aaaaaaaaaa 832

55 <210> 49
 <211> 831
 <212> DNA
 <213> Homo sapiens

60 <220>
 <221> CDS
 <222> 234..491

```

<220>
<221> sig_peptide
<222> 234..293
<223> Von Heijne matrix
5   score 4.85037394589162
    seq AVAGLPALGFTGA/GI

<400> 49
10 gtttactcgc tgctgtgccc atctatcagc aggctccggg ctgaagattg cttctcttct      60
   ctcctccaag gtctagtgac ggagcccgcg cgcggcgcca ccatgcggca gaaggcgta      120
   tcgcttttct ctgctacctg ctgctcttca cttgcagtgg ggtggaggca ggtaagaaaa      180
   agtgctcgga gagctcggac agcggctccg ggttctggaa ggccctgacc ttc atg      236
                                   Met
                                   -20
15 gcc gtc gga gga gga ctc gca gtc gcc ggg ctg ccc gcg ctg ggc ttc      284
   Ala Val Gly Gly Gly Leu Ala Val Ala Gly Leu Pro Ala Leu Gly Phe
                                   -15                                   -10                                   -5
   acc ggc gcc ggc atc gcg gcc aac tcg gtg gct gcc tcg ctg atg agc      332
   Thr Gly Ala Gly Ile Ala Ala Asn Ser Val Ala Ala Ser Leu Met Ser
20   tgg tct gcg atc ctg aat ggg ggc ggc gtg ccc gcc ggg ggg cta gtg      380
   Trp Ser Ala Ile Leu Asn Gly Gly Gly Val Pro Ala Gly Gly Leu Val
   15                                   20                                   25
   gcc acg ctg cag agc ctc ggg gct ggt ggc agc agc gtc gtc ata ggt      428
25   Ala Thr Leu Gln Ser Leu Gly Ala Gly Gly Ser Ser Val Val Ile Gly
   30                                   35                                   40                                   45
   aat att ggt gcc ctg atg ggc tac gcc acc cac aag tat ctc gat agt      476
   Asn Ile Gly Ala Leu Met Gly Tyr Ala Thr His Lys Tyr Leu Asp Ser
   50                                   55                                   60
30   gag gag gat gag gag tagccagcag ctcccagaac ctcttcttcc ttcttggcct      531
   Glu Glu Asp Glu Glu
   65
   aactcttcca gttaggatct agaactttgc cttttttttt tttttttttt tttttttgag      591
   atgggttctc actatattgt ccaggctaga gtgcagtggc tattcacaga tgcgaacata      651
35   gtacactgca gcctccaact cctagcctca agtgatcctc ctgtctcaac ctcccaagta      711
   ggattacaag catgcgcoga cgatgccag aatccagaac tttgtctatc actctcccca      771
   acaacctaga tgtgaaaaca gaataaactt caccagaaa gcaaaaaaaaa aaaaaaaaaa      831

<210> 50
40 <211> 917
   <212> DNA
   <213> Homo sapiens

<220>
45 <221> CDS
   <222> 180..800

<220>
<221> sig_peptide
50 <222> 180..248
   <223> Von Heijne matrix
       score 14.6828672385356
       seq ILLLLWLIAPSRA/CT

55 <400> 50
   acccttgget tctgcaactga tgggtgggtgg atgagtaatg catccaggaa gcctggaggc      60
   ctgtggtttc cgcacccgct gccacccccg cccctagcgt ggacatttat cctctagcgc      120
   tcaggccctg ccgccatcgc cgcagatcca gcgccagag agacaccaga gaaccacc      179
   atg gcc ccc ttt gag ccc ctg gct tct ggc atc ctg ttg ttg ctg tgg      227
60   Met Ala Pro Phe Glu Pro Leu Ala Ser Gly Ile Leu Leu Leu Leu Trp
   -20                                   -15                                   -10
   ctg ata gcc ccc agc agg gcc tgc acc tgt gtc cca ccc cac cca cag      275
   Leu Ile Ala Pro Ser Arg Ala Cys Thr Cys Val Pro Pro His Pro Gln

```

```

      -5              1              5
acg gcc ttc tgc aat tcc gac ctc gtc atc agg gcc aag ttc gtg ggg      323
Thr Ala Phe Cys Asn Ser Asp Leu Val Ile Arg Ala Lys Phe Val Gly
10              15              20              25
5  aca cca gaa gtc aac cag acc acc tta tac cag cgt tat gag atc aag      371
Thr Pro Glu Val Asn Gln Thr Thr Leu Tyr Gln Arg Tyr Glu Ile Lys
      30              35              40
atg acc aag atg tat aaa ggg ttc caa gcc tta ggg gat gcc gct gac      419
Met Thr Lys Met Tyr Lys Gly Phe Gln Ala Leu Gly Asp Ala Ala Asp
10      45              50              55
atc cgg ttc gtc tac acc ccc gcc atg gag agt gtc tgc gga tac ttc      467
Ile Arg Phe Val Tyr Thr Pro Ala Met Glu Ser Val Cys Gly Tyr Phe
      60              65              70
cac agg tcc cac aac cgc agc gag gag ttt ctc att gct gga aaa ctg      515
15  His Arg Ser His Asn Arg Ser Glu Glu Phe Leu Ile Ala Gly Lys Leu
      75              80              85
cag gat gga ctc ttg cac atc act acc tgc agt ttt gtg gct ccc tgg      563
Gln Asp Gly Leu Leu His Ile Thr Thr Cys Ser Phe Val Ala Pro Trp
20      90              95              100
aac agc ctg agc tta gct cag cgc cgg ggc ttc acc aag acc tac act      611
Asn Ser Leu Ser Leu Ala Gln Arg Arg Gly Phe Thr Lys Thr Tyr Thr
      110              115              120
gtt ggc tgt gag gaa tgc aca gtg ttt ccc tgt tta tcc ttc ccc tgc      659
Val Gly Cys Glu Glu Cys Thr Val Phe Pro Cys Leu Ser Phe Pro Cys
25      125              130              135
aaa ctg cag agt ggc act cat tgc ttg tgg acg gac cag ctc ctc caa      707
Lys Leu Gln Ser Gly Thr His Cys Leu Trp Thr Asp Gln Leu Leu Gln
      140              145              150
ggc tct gaa aag ggc ttc cag tcc cgt cac ctt gcc tgc ctg cct cgg      755
30  Gly Ser Glu Lys Gly Phe Gln Ser Arg His Leu Ala Cys Leu Pro Arg
      155              160              165
gag cca ggg ctg tgc acc tgg cag tcc ctg cgg tcc cag ata gcc      800
Glu Pro Gly Leu Cys Thr Trp Gln Ser Leu Arg Ser Gln Ile Ala
170              175              180
35  tgaatcctgc ccggagtgga agctgaagcc tgcacagtgt ccaccctgtt cccactccca      860
tctttcttcc ggacaatgaa ataaagagtt accaccagc aaaaaaaaaa aaaaaaa      917

<210> 51
<211> 621
40 <212> DNA
    <213> Homo sapiens

<220>
<221> CDS
45 <222> 140..472

<220>
<221> sig_peptide
<222> 140..211
50 <223> Von Heijne matrix
    score 8.44884907465122
    seq FVVFSLFILICAMA/GD

<400> 51
55 attttttttt catatctgac atttctatgt cctatgacgg tttcacagct atcctacttt      60
   ggagaagatg ctggaaattc agagtttccg ccagagaata tatgcctgaa ctaaaagagg      120
   aagtgggtcta taggagaaa atg aaa tat gat tgt ccc ttc agt ggg aca tca      172
                        Met Lys Tyr Asp Cys Pro Phe Ser Gly Thr Ser
                        -20              -15
60 ttt gtg gtc ttc tct ctc ttt ttg atc tgt gca atg gct gga gat gta      220
   Phe Val Val Phe Ser Leu Phe Leu Ile Cys Ala Met Ala Gly Asp Val
                        -10              -5              1
   gtc tac gct gac atc aaa act gtt cgg act tcc ccg tta gaa ctc gcg      268

```

```

Val Tyr Ala Asp Ile Lys Thr Val Arg Thr Ser Pro Leu Glu Leu Ala
  5          10          15
ttt cca ctt cag aga tct gtt tct ttc aac ttt tct act gtc cat aaa      316
Phe Pro Leu Gln Arg Ser Val Ser Phe Asn Phe Ser Thr Val His Lys
5 20          25          30          35
tca tgt cct gcc aaa gac tgg aag gtg cat aag gga aaa tgt tac tgg      364
Ser Cys Pro Ala Lys Asp Trp Lys Val His Lys Gly Lys Cys Tyr Trp
          40          45          50
att gct gaa act aag aaa tct tgg aac aaa agt caa aat gac tgt gcc      412
10 Ile Ala Glu Thr Lys Lys Ser Trp Asn Lys Ser Gln Asn Asp Cys Ala
          55          60          65
ata aac aat tca tat ctc atg gtg att caa gac att act gct atg gtg      460
Ile Asn Asn Ser Tyr Leu Met Val Ile Gln Asp Ile Thr Ala Met Val
          70          75          80
15 aga ttt aac att tagaggtgac agcatccccc acactggcag ttaatttttt      512
Arg Phe Asn Ile
          85
gtctacaaac ttggcaaaag tctgtgaaaa gaagtttcaa cttcatgtgt tattaactat      572
acaaatatta gttgaatgaa ttgttgaaatt acaaaaaaaaa aaaaaaaaaa      621
20
<210> 52
<211> 673
<212> DNA
<213> Homo sapiens
25
<220>
<221> CDS
<222> 68..484
30
<220>
<221> sig_peptide
<222> 68..112
<223> Von Heijne matrix
      score 4.93618539864455
35      seq AVVFVFSLLDCCA/LI

<400> 52
ctatcagggg tgggtcgggg catccgagcg ggtttgacgg aaggagcggc ggcgacggag      60
gaggagg atg gag gcg gtg gtg ttc gtc ttc tct ctc ctc gat tgt tgc      109
40      Met Glu Ala Val Val Phe Val Phe Ser Leu Leu Asp Cys Cys
          -15          -10          -5
gcg ctc atc ttc ctc tcg gtc tac ttc ata att aca ttg tct gat tta      157
Ala Leu Ile Phe Leu Ser Val Tyr Phe Ile Ile Thr Leu Ser Asp Leu
          1          5          10          15
45 gaa tgt gat tac att aat gct aga tca tgt tgc tca aaa tta aac aag      205
Glu Cys Asp Tyr Ile Asn Ala Arg Ser Cys Cys Ser Lys Leu Asn Lys
          20          25          30
tggtgtg att cca gaa ttg att ggc cat acc att gtc act gta tta ctg      253
Trp Val Ile Pro Glu Leu Ile Gly His Thr Ile Val Thr Val Leu Leu
          35          40          45
50 ctc atg tca ttg cac tgg ttc atc ttc ctt ctc aac tta cct gtt gcc      301
Leu Met Ser Leu His Trp Phe Ile Phe Leu Leu Asn Leu Pro Val Ala
          50          55          60
act tgg aat ata tat cga tac att atg gtg ccg agt ggt aac atg gga      349
55 Thr Trp Asn Ile Tyr Arg Tyr Ile Met Val Pro Ser Gly Asn Met Gly
          65          70          75
gtg ttt gat cca aca gaa ata cac aat cga ggg cag ctg aag tca cac      397
Val Phe Asp Pro Thr Glu Ile His Asn Arg Gly Gln Leu Lys Ser His
          80          85          90          95
60 atg aaa gaa gcc atg atc aag ctt ggt ttc cac ttg ctc tgc ttc ttc      445
Met Lys Glu Ala Met Ile Lys Leu Gly Phe His Leu Leu Cys Phe Phe
          100          105          110
atg tat ctt tat agt atg atc tta gct ttg ata aat gac tgaagctgga      494

```

Met Tyr Leu Tyr Ser Met Ile Leu Ala Leu Ile Asn Asp
 115 120
 gaagccgtgg ttgaagtcag cctacactac agtgcacagt tgaggagcca gagacttctt 554
 aaatcatcct tagaaccgtg accatagcag tatatatattt cctcttggaa caaaaaacta 614
 5 tttttgctgt atttttacca tataaagtat ttaaaaaaca cgaaaaaaaa aaaaaaaaaa 673

<210> 53
 <211> 897
 <212> DNA
 10 <213> Homo sapiens

<220>
 <221> CDS
 <222> 38..517
 15 <220>
 <221> sig_peptide
 <222> 38..118
 <223> Von Heijne matrix
 20 score 7.20400999800742
 seq VLWLSGLSEPGAA/RQ

<400> 53
 agattggggac agtcgccagg gatggctgag cgtgaag atg cag cgg gtg tcc ggg 55
 25 Met Gln Arg Val Ser Gly
 -25
 ctg ctc tcc tgg acg ctg agc aga gtc ctg tgg ctc tcc ggc ctc tct 103
 Leu Leu Ser Trp Thr Leu Ser Arg Val Leu Trp Leu Ser Gly Leu Ser
 -20 -15 -10
 30 gag ccg gga gct gcc cgg cag ccc cgg atc atg gaa gag aaa gcg cta 151
 Glu Pro Gly Ala Ala Arg Gln Pro Arg Ile Met Glu Glu Lys Ala Leu
 -5 1 5 10
 gag gtt tat gat ttg att aga act atc cgg gac cca gaa aag ccc aat 199
 Glu Val Tyr Asp Leu Ile Arg Thr Ile Arg Asp Pro Glu Lys Pro Asn
 35 15 20 25
 act tta gaa gaa ctg gaa gtg gtc tcg gaa agt tgt gtg gaa gtt cag 247
 Thr Leu Glu Glu Leu Glu Val Ser Glu Ser Cys Val Glu Val Gln
 30 35 40
 gag ata aat gaa gaa gaa tat ctg gtt att atc agg ttc acg cca aca 295
 40 Glu Ile Asn Glu Glu Glu Tyr Leu Val Ile Ile Arg Phe Thr Pro Thr
 45 50 55
 gta cct cat tgc tct ttg gcg act ctt att ggg ctg tgc tta aga gta 343
 Val Pro His Cys Ser Leu Ala Thr Leu Ile Gly Leu Cys Leu Arg Val
 60 65 70 75
 45 aaa ctt cag cga tgt tta cca ttt aaa cat aag ttg gaa atc tac att 391
 Lys Leu Gln Arg Cys Leu Pro Phe Lys His Lys Leu Glu Ile Tyr Ile
 80 85 90
 tct gaa gga acc cac tca aca gaa gaa gac atc aat aag cag ata aat 439
 Ser Glu Gly Thr His Ser Thr Glu Glu Asp Ile Asn Lys Gln Ile Asn
 50 95 100 105
 gac aaa gag cga gtg gca gct gca atg gaa aac ccc aac tta cgg gaa 487
 Asp Lys Glu Arg Val Ala Ala Ala Met Glu Asn Pro Asn Leu Arg Glu
 110 115 120
 att gtg gaa cag tgt gtc ctt gaa cct gac tgatagctgt tttaagagcc 537
 55 Ile Val Glu Gln Cys Val Leu Glu Pro Asp
 125 130
 actggcctgt aattgtttga tatattttgta actcttttga taatgtcaga gactcatgtt 597
 taatacatag gtgattttgta cctcagagca ttttttaaag gattctttcc aagcgagatt 657
 taattataag gtagtaccta atttgttcaa tgtataacat tctcaggatt tgtaacactt 717
 60 aaatgatcag acagaataat attttctagt tattatgtgt aagatgagtt gctatttttc 777
 tgatgctcat tctgatacaa ctatttttcg tgtcaaatat ctactgtgcc caaatgtact 837
 caattttaaatt cattactctg taaaataaat aagcagatga ttcttataaa aaaaaaaaaa 897


```

<210> 54
<211> 1101
<212> DNA
<213> Homo sapiens
5
<220>
<221> CDS
<222> 92..634

10 <220>
<221> sig_peptide
<222> 92..139
<223> Von Heijne matrix
      score 7.36306712986597
15      seq FLLLTCLFITGTS/VS

<400> 54
cttaaaaaaaaa aaagtgcttg aaagagaagg ggacaaagga acaccagtat taagaggatt      60
ttccagtgtt tctggcagtt ggtccagaag g atg cct cca ttc ctg ctt ctc      112
20                                     Met Pro Pro Phe Leu Leu Leu
                                     -15                                     -10
acc tgc ctc ttc atc aca ggc acc tcc gtg tca ccc gtg gcc cta gat      160
Thr Cys Leu Phe Ile Thr Gly Thr Ser Val Ser Pro Val Ala Leu Asp
                                     -5                                     1                                     5
25 cct tgt tct gct tac atc agc ctg aat gag ccc tgg agg aac act gac      208
Pro Cys Ser Ala Tyr Ile Ser Leu Asn Glu Pro Trp Arg Asn Thr Asp
                                     10                                     15                                     20
cac cag ttg gat gag tct caa ggt cct cct cta tgt gac aac cat gtg      256
His Gln Leu Asp Glu Ser Gln Gly Pro Pro Leu Cys Asp Asn His Val
30      25      30      35
aat ggg gag tgg tac cac ttc acg ggc atg gcg gga gat gcc atg cct      304
Asn Gly Glu Trp Tyr His Phe Thr Gly Met Ala Gly Asp Ala Met Pro
40      40      45      50      55
acc ttc tgc ata cca gaa aac cac tgt gga acc cac gca cct gtc tgg      352
Thr Phe Cys Ile Pro Glu Asn His Cys Gly Thr His Ala Pro Val Trp
35      60      65
ctc aat ggc agc cac ccc cta gaa ggc gac ggc att gtg caa cgc cag      400
Leu Asn Gly Ser His Pro Leu Glu Gly Asp Gly Ile Val Gln Arg Gln
75      80      85
40 gct tgt gcc agc ttc aat ggg aac tgc tgt ctc tgg aac acc acg gtg      448
Ala Cys Ala Ser Phe Asn Gly Asn Cys Cys Leu Trp Asn Thr Thr Val
90      95      100
gaa gtc aag gct tgc cct gga ggc tac tat gtg tat cgt ctg acc aag      496
Glu Val Lys Ala Cys Pro Gly Gly Tyr Tyr Val Tyr Arg Leu Thr Lys
45      105      110      115
ccc agc gtc tgc ttc cac gtc tac tgt ggt cgt gag tac ctt ccc tgt      544
Pro Ser Val Cys Phe His Val Tyr Cys Gly Arg Glu Tyr Leu Pro Cys
120      125      130      135
gct ctt ttt ctc cac caa caa ggc cac agg tgg agt cca aaa gtg ccc      592
Ala Leu Phe Leu His Gln Gln Gly His Arg Trp Ser Pro Lys Val Pro
50      140      145      150
aat tat agg ata tgc agt tac agt ggc aac tat atc tca atc      634
Asn Tyr Arg Ile Cys Ser Tyr Ser Gly Asn Tyr Ile Ser Ile
155      160      165
55 tgaacaacat tgatgtgggg ctaaagatac tctgatttct gagatctctt cttagaactt      694
ctgaaaaaatt cctgaagaaa tagaagggga aaggagctat gactttgatc agttcttttt      754
aattttgtct gaattccatt caaacaaaac attagaaaat gaaacattgg gccaggcgca      814
gtggctcatg cctgtaatcc cagcactttg ggaggctgag gcgggtggat cacaagatca      874
ggagttaag accagcctgg ccaatatggt gaaaccctgt ctctactaga aatacaaaaa      934
60 ttagacaggc gtggtggcag gcaactgtaa cccagctac cggggaggct gaggcaggag      994
aattgcttga acccgggagg tggacgttgc ggtcaggcga aaatcgtgcc attgcactcc      1054
agcctgggtg acagagtgag actctgattc aaaaaaaaaa aaaaaaa      1101

```

<210> 55
 <211> 1047
 <212> DNA
 <213> Homo sapiens
 5
 <220>
 <221> CDS
 <222> 27..767
 10 <220>
 <221> sig_peptide
 <222> 27..80
 <223> Von Heijne matrix
 score 8.96664802487992
 15 seq LFCLAVLAASSFS/KA
 <400> 55
 agcagaggcc ctacacccac cgaggc atg ggg ctc cct ggg ctg ttc tgc ttg 53
 Met Gly Leu Pro Gly Leu Phe Cys Leu
 20
 gcc gtg ctg gct gcc agc agc ttc tcc aag gca cgg gag gaa gaa att 101
 Ala Val Leu Ala Ala Ser Ser Phe Ser Lys Ala Arg Glu Glu Glu Ile
 -5 1 5
 acc cct gtg gtc tcc att gcc tac aaa gtc ctg gaa gtt ttc ccc aaa 149
 25 Thr Pro Val Val Ser Ile Ala Tyr Lys Val Leu Glu Val Phe Pro Lys
 10 15 20
 ggc cgc tgg gtg ctc ata acc tgc tgt gca ccc cag cca cca ccg ccc 197
 Gly Arg Trp Val Leu Ile Thr Cys Cys Ala Pro Gln Pro Pro Pro Pro
 25 30 35
 30 atc acc tat tcc ctc tgt gga acc aag aac atc aag gtg gcc aag aag 245
 Ile Thr Tyr Ser Leu Cys Gly Thr Lys Asn Ile Lys Val Ala Lys Lys
 40 45 50 55
 gtg gtg aag acc cac gag ccg gcc tcc ttc aac ctc aac gtc aca ctc 293
 Val Val Lys Thr His Glu Pro Ala Ser Phe Asn Leu Asn Val Thr Leu
 35 60 65 70
 aag tcc agt cca gac ctg ctc acc tac ttc tgc cgg gcg tcc tcc acc 341
 Lys Ser Ser Pro Asp Leu Leu Thr Tyr Phe Cys Arg Ala Ser Ser Thr
 75 80 85
 tca ggt gcc cat gtg gac agt gcc agg cta cag atg cac tgg gag ctg 389
 40 Ser Gly Ala His Val Asp Ser Ala Arg Leu Gln Met His Trp Glu Leu
 90 95 100
 tgg tcc aag cca gtg tct gag ctg ccg gcc aac ttc act ctg cag gac 437
 Trp Ser Lys Pro Val Ser Glu Leu Arg Ala Asn Phe Thr Leu Gln Asp
 105 110 115
 45 aga ggg gca ggc ccc agg gtg gag atg atc tgc cag gcg tcc tcg ggc 485
 Arg Gly Ala Gly Pro Arg Val Glu Met Ile Cys Gln Ala Ser Ser Gly
 120 125 130 135
 agc cca cct atc acc aac agc ctg atc ggg aag gat ggg cag gtc cac 533
 Ser Pro Pro Ile Thr Asn Ser Leu Ile Gly Lys Asp Gly Gln Val His
 140 145 150
 50 ctg cag cag aga cca tgc cac agg cag cct gcc aac ttc tcc ttc ctg 581
 Leu Gln Gln Arg Pro Cys His Arg Gln Pro Ala Asn Phe Ser Phe Leu
 155 160 165
 ccg agc cag aca tcg gac tgg ttc tgg tgc cag gct gca aac aac gcc 629
 55 Pro Ser Gln Thr Ser Asp Trp Phe Trp Cys Gln Ala Ala Asn Asn Ala
 170 175 180
 aat gtc cag cac agc gcc ctc aca gtg gtg ccc cca gga ggg ttg ccc 677
 Asn Val Gln His Ser Ala Leu Thr Val Val Pro Pro Gly Gly Leu Pro
 185 190 195
 60 agg gca ccc acc atc gtg ctg gtt ggc agc ctt gcc tcc act gcg gcc 725
 Arg Ala Pro Thr Ile Val Leu Val Gly Ser Leu Ala Ser Thr Ala Ala
 200 205 210 215
 atc acc tcc agg atg ctg ggc tgg acc acg tgg gcc agg tgg 767

Ile Thr Ser Arg Met Leu Gly Trp Thr Thr Trp Ala Arg Trp
 220 225

tgaccagaag atggaggact ggcaggggtcc cctggagagc cccatccttg ccttgccgct 827
 ctacaggagc acccgccgtc tgagtgaaga ggagtttggg gggttcagga tagggaatgg 887
 5 ggaggtcaga ggacgcaaag cagcagccat gtagaatgaa ccgtccagag agccaagcac 947
 ggcagaggac tgcaggccat cagcgtgcac tggtcgtatt tggagttcat gcaaaatgag 1007
 tgtgttttag ctgctcttgc cacaaaaaaa aaaaaaaaaa 1047

<210> 56
 10 <211> 925
 <212> DNA
 <213> Homo sapiens

<220>
 15 <221> CDS
 <222> 4..399

<220>
 <221> sig_peptide
 20 <222> 4..126
 <223> Von Heijne matrix
 score 4.34454795165846
 seq RVVSWLFSIVVFG/SI

25 <400> 56

acg atg gaa ggg ggt gcg tac gga gcg ggc aaa gcc ggg ggc gcc ttc 48
 Met Glu Gly Gly Ala Tyr Gly Ala Gly Lys Ala Gly Gly Ala Phe
 -40 -35 -30

gac ccc tac acc ctg gtc cgg cag ccg cac acc atc ctg cgc gtc gtg 96
 30 Asp Pro Tyr Thr Leu Val Arg Gln Pro His Thr Ile Leu Arg Val Val
 -25 -20 -15

tct tgg ctg ttc tcc ata gtg gtg ttc ggc tcc atc gtg aac gag ggc 144
 Ser Trp Leu Phe Ser Ile Val Val Phe Gly Ser Ile Val Asn Glu Gly
 -10 -5 1 5

tac ctc aac agc gcc tcc gag ggg gag cag ttc tgc atc tac aac cgc 192
 35 Tyr Leu Asn Ser Ala Ser Glu Gly Glu Gln Phe Cys Ile Tyr Asn Arg
 10 15 20

aac ccc aac gcc tgc agc tat ggc gtg gcc gtg ggc gtg ctc gcc ttc 240
 Asn Pro Asn Ala Cys Ser Tyr Gly Val Ala Val Gly Val Leu Ala Phe
 25 30 35

ctc acc tgc ctg ctg tac ctg gcc ctg gac gtg tac ttc ccg cag atc 288
 40 Leu Thr Cys Leu Leu Tyr Leu Ala Leu Asp Val Tyr Phe Pro Gln Ile
 40 45 50

agc agc gtc aag gac cgc aag aaa gcc gtc ctg tcc gac atc ggt gtc 336
 45 Ser Ser Val Lys Asp Arg Lys Lys Ala Val Leu Ser Asp Ile Gly Val
 55 60 65 70

tcg ggt gag ccc cac cca gca ggt acc ccc tgc aca gag tct aca gag 384
 Ser Gly Glu Pro His Pro Ala Gly Thr Pro Cys Thr Glu Ser Thr Glu
 75 80 85

ggc tgt ccc ggg cca taggaggcgg ctgccaccct tcttcccatg ttccagatga 439
 50 Gly Cys Pro Gly Pro
 90

gggaaatgag ccttctgggc tttcctctgg ttcgtgggat tctgctacct ggccaaccag 499
 tggcaggtct ccaagcccaa ggacaacca ctgaacgaag ggacggacgc agcccgggccc 559
 55 gccatgcct tctccttttt ctccatcttc acctggagcc tgaccgcagc cctggccgtg 619
 cggagattca aggacctaa gcttcaggag gactacagca cactgttccc tgcttcggca 679
 cagccgtagg cctccccggc ttgcagaggc cggcagccct gtatcaccct tggcagtgag 739
 gtggcaggag cagcctagt ccagaaatgt ccaagatgcc agggcatgca gggcagtgga 799
 aggctggctt gaggaaccaa ttcaggttct ccaactgactc attcattcct tcaccgcctc 859
 60 cttcattgat tcttcattgct ttcattcatt cagtaaacad ttattgagta aaaaaaaaaa 919
 aaaaaa 925

<210> 57

<211> 1240
 <212> DNA
 <213> Homo sapiens

5 <220>
 <221> CDS
 <222> 127..879

<220>
 10 <221> sig_peptide
 <222> 127..198
 <223> Von Heijne matrix
 score 5.38660866264012
 seq ALCSVCSMSVLRA/YP

15 <400> 57
 agtctaggat cctcacacca gctacttgca agggagaagg aaaaggccag taaggcctgg 60
 gccaggagag tcccgcacagg agtgtcaggt ttcaatctca gcaccagcca ctcagagcag 120
 ggcacg atg ttg ggg gcc cgc ctc agg ctc tgg gtc tgt gcc ttg tgc 168
 20 Met Leu Gly Ala Arg Leu Arg Leu Trp Val Cys Ala Leu Cys
 -20 -15
 agc gtc tgc agc atg agc gtc ctc aga gcc tat ccc aat gcc tcc cca 216
 Ser Val Cys Ser Met Ser Val Leu Arg Ala Tyr Pro Asn Ala Ser Pro
 -10 -5 1 5
 25 ctg ctc ggc tcc agc tgg ggt ggc ctg atc cac ctg tac aca gcc aca 264
 Leu Leu Gly Ser Ser Trp Gly Gly Leu Ile His Leu Tyr Thr Ala Thr
 10 15 20
 gcc agg aac agc tac cac ctg cag atc cac aag aat ggc cat gtg gat 312
 Ala Arg Asn Ser Tyr His Leu Gln Ile His Lys Asn Gly His Val Asp
 25 30 35
 30 ggc gca ccc cat cag acc atc tac agt gcc ctg atg atc aga tca gag 360
 Gly Ala Pro His Gln Thr Ile Tyr Ser Ala Leu Met Ile Arg Ser Glu
 40 45 50
 gat gct ggc ttt gtg gtg att aca ggt gtg atg agc aga aga tac ctc 408
 35 Asp Ala Gly Phe Val Val Ile Thr Gly Val Met Ser Arg Arg Tyr Leu
 55 60 65 70
 tgc atg gat ttc aga ggc aac att ttt gga tca cac tat ttc gac ccg 456
 Cys Met Asp Phe Arg Gly Asn Ile Phe Gly Ser His Tyr Phe Asp Pro
 75 80 85
 40 gag aac tgc agg ttc caa cac cag acg ctg gaa aac ggg tac gac gtc 504
 Glu Asn Cys Arg Phe Gln His Gln Thr Leu Glu Asn Gly Tyr Asp Val
 90 95 100
 tac cac tct cct cag tat cac ttc ctg gtc agt ctg ggc cgg gcg aag 552
 Tyr His Ser Pro Gln Tyr His Phe Leu Val Ser Leu Gly Arg Ala Lys
 105 110 115
 45 aga gcc ttc ctg cca ggc atg aac cca ccc ccg tac tcc cag ttc ctg 600
 Arg Ala Phe Leu Pro Gly Met Asn Pro Pro Pro Tyr Ser Gln Phe Leu
 120 125 130
 tcc cgg agg aac gag atc ccc cta att cac ttc aac acc ccc ata cca 648
 50 Ser Arg Arg Asn Glu Ile Pro Leu Ile His Phe Asn Thr Pro Ile Pro
 135 140 145 150
 cgg cgg cac acc cgg agc gcc gag gac gac tgc gag cgg gac ccc ctg 696
 Arg Arg His Thr Arg Ser Ala Glu Asp Asp Ser Glu Arg Asp Pro Leu
 155 160 165
 55 aac gtg ctg aag ccc cgg gcc cgg atg acc ccg gcc ccg gcc tcc tgt 744
 Asn Val Leu Lys Pro Arg Ala Arg Met Thr Pro Ala Pro Ala Ser Cys
 170 175 180
 tca cag gag ctc ccg agc gcc gag gac aac agc ccg atg gcc agt gac 792
 Ser Gln Glu Leu Pro Ser Ala Glu Asp Asn Ser Pro Met Ala Ser Asp
 185 190 195
 60 cca tta ggg gtg gtc agg ggc ggt cga gtg aac acg cac gct ggg gga 840
 Pro Leu Gly Val Val Arg Gly Gly Arg Val Asn Thr His Ala Gly Gly
 200 205 210

acg ggc ccg gaa ggc tgc cgc ccc ttc gcc aag ttc atc taggggtcgct 889
 Thr Gly Pro Glu Gly Cys Arg Pro Phe Ala Lys Phe Ile
 215 220 225

5 ggaaggggcac cctctttaac ccattccctca gcaaacgcag ctcttcccaa ggaccagggtc 949
 ccttgacgtt ccgaggatgg gaaagggtgac aggggcatgt atggaatttg ctgcttctct 1009
 ggggtccctt ccacaggagg tctgtgaga accaaccttt gagggccaag tcatgggggtt 1069
 tcaccgcctt cctcactcca tatagaacac ctttcccaat aggaaacccc aacagggtaaa 1129
 ctagaaattt ccccttcattg aaggtagaga gaagggtgt ctcccaacat atttctcttc 1189
 cttgtgcctc tcctctttat cactttttaag catgaaaaaa aaaaaaaaaa a 1240

10 <210> 58
 <211> 902
 <212> DNA
 <213> Homo sapiens

15 <220>
 <221> CDS
 <222> 156..566

20 <220>
 <221> sig_peptide
 <222> 156..221
 <223> Von Heijne matrix
 score 5.67458379966095

25 seq LVSMAGRVCLCQG/SA

<400> 58
 atttcccagc gtgcctcagg aagggcgcca ggactgcatt ttgctccgga gcgtccagag 60
 tcctggccct gagcgggaat cgcagtggcc gaggtgagc ggcaggcgga tcgccccgac 120
 30 cctcactcct ggcgtctgag tctctggcgt agccc atg ctg agt ggg cgg ctg 173
 Met Leu Ser Gly Arg Leu
 -20

gtc ctg ggt ctg gtc tcc atg gct ggc cgc gtt tgt ttg tgc cag ggc 221
 Val Leu Gly Leu Val Ser Met Ala Gly Arg Val Cys Leu Cys Gln Gly
 35 -15 -10 -5
 agc gcg gga tcc ggg gcc atc ggt ccg gtg gag gcc gcc att cgc acg 269
 Ser Ala Gly Ser Gly Ala Ile Gly Pro Val Glu Ala Ala Ile Arg Thr
 1 5 10 15

40 aag ttg gag gag gcc ctg agc ccc gag gtg cta gag ctt cgc aac gag 317
 Lys Leu Glu Glu Ala Leu Ser Pro Glu Val Leu Glu Leu Arg Asn Glu
 20 25 30

agc ggt ggc cac gcg gtc ccg cca ggc agt gag act cac ttc cgc gtg 365
 Ser Gly Gly His Ala Val Pro Pro Gly Ser Glu Thr His Phe Arg Val
 35 40 45

45 gct gtg gtg agc tct cgt ttc gag gga ctg agc ccc cta caa cga cac 413
 Ala Val Val Ser Ser Arg Phe Glu Gly Leu Ser Pro Leu Gln Arg His
 50 55 60

cgg ctg gtc cac gca gcg ctg gcc gag gag ctg gga ggt ccg gtc cat 461
 Arg Leu Val His Ala Ala Leu Ala Glu Glu Leu Gly Gly Pro Val His
 50 65 70 75 80

gcg ctg gcc atc cag gca cgg acc ccc gcc cag tgg aga gag aac tct 509
 Ala Leu Ala Ile Gln Ala Arg Thr Pro Ala Gln Trp Arg Glu Asn Ser
 85 90 95

cag ctg gac act agc ccc cca tgc ctg ggt ggg aac aag aaa act cta 557
 55 Gln Leu Asp Thr Ser Pro Pro Cys Leu Gly Gly Asn Lys Lys Thr Leu
 100 105 110

gga acc ccc tgaaccccaa gagagggagg accaggatcc gaatgggctg 606
 Gly Thr Pro
 115

60 ggtgagcacg aattaccgag gccttccctt tgatacagtc caggatttgt aagggatgaa 666
 gacccctggg cccattctg ttgggggtcca tacatactct ccgaagatag caacttgctt 726
 caggtcaaag tgaacccgag aaaagagaag aatcactcac tactgtctt gccttgact 786
 attcaggaag ggcagcccgg atgttccatg ttaaactcgtg acagaattgc accagacctg 846

atgagttgga aacaatccta tacattaaaa gaaattacac taaaaaaaaa aaaaaa 902

<210> 59

<211> 1969

5 <212> DNA

<213> Homo sapiens

<220>

<221> CDS

10 <222> 35..1657

<220>

<221> sig_peptide

<222> 35..118

15 <223> Von Heijne matrix

score 3.75144398608723

seq SGLLLQVLFRLIT/FV

<400> 59

20 atttttcttg gtgtctgagc ctggcgcgga ggct atg ggc agc cag gag gtg ctg 55
Met Gly Ser Gln Glu Val Leu

-25

ggc cac gcg gcc cgg ctg gcc tcc tcc ggt ctc ctc ctg cag gtg ttg 103
Gly His Ala Ala Arg Leu Ala Ser Ser Gly Leu Leu Leu Gln Val Leu

25 -20 -15 -10
ttt cgg ttg atc acc ttt gtc ttg aat gca ttt att ctt cgc ttc ctg 151
Phe Arg Leu Ile Thr Phe Val Leu Asn Ala Phe Ile Leu Arg Phe Leu

-5 1 5 10
tca aag gaa atc gtt ggc gta gta aat gta aga cta acg ctg ctt tac 199
Ser Lys Glu Ile Val Gly Val Val Asn Val Arg Leu Thr Leu Leu Tyr

30 15 20 25
tca acc acc ctc ttc ctg gcc aga gag gcc ttc cgc aga gca tgt ctc 247
Ser Thr Thr Leu Phe Leu Ala Arg Glu Ala Phe Arg Arg Ala Cys Leu

35 30 35 40
agt ggg ggc acc cag cga gac tgg agc cag acc ctc aac ctg ctg tgg 295
Ser Gly Gly Thr Gln Arg Asp Trp Ser Gln Thr Leu Asn Leu Leu Trp

45 45 50 55
cta aca gtc ccc ctg ggt gtg ttt tgg tcc tta ttc ctg ggc tgg atc 343
Leu Thr Val Pro Leu Gly Val Phe Trp Ser Leu Phe Leu Gly Trp Ile

40 60 65 70 75
tgg ttg cag ctg ctt gaa gtg cct gat cct aat gtt gtc cct cac tat 391
Trp Leu Gln Leu Leu Glu Val Pro Asp Pro Asn Val Val Pro His Tyr

80 85 90
gca act gga gtg gtg ctg ttt ggt ctc tcg gca gtg gtg gag ctt cta 439
Ala Thr Gly Val Val Leu Phe Gly Leu Ser Ala Val Val Glu Leu Leu

95 100 105
gga gag ccc ttt tgg gtc ttg gca caa gca cat atg ttt gtg aag ctc 487
Gly Glu Pro Phe Trp Val Leu Ala Gln Ala His Met Phe Val Lys Leu

110 115 120
aag gtg att gca gag agc ctg tcg gta att ctt aag agc gtt ctg aca 535
Lys Val Ile Ala Glu Ser Leu Ser Val Ile Leu Lys Ser Val Leu Thr

125 130 135
gct ttt ctc gtg ctg tgg ttg cct cac tgg gga ttg tac att ttc tct 583
Ala Phe Leu Val Leu Trp Leu Pro His Trp Gly Leu Tyr Ile Phe Ser

55 140 145 150 155
ttg gcc cag ctt ttc tat acc aca gtt ctg gtg ctc tgc tat gtt att 631
Leu Ala Gln Leu Phe Tyr Thr Thr Val Leu Val Leu Cys Tyr Val Ile

160 165 170
tat ttc aca aag tta ctg ggt tcc cca gaa tca acc aag ctt caa act 679
Tyr Phe Thr Lys Leu Leu Gly Ser Pro Glu Ser Thr Lys Leu Gln Thr

175 180 185
ctt cct gtc tcc aga ata aca gat ctg tta ccc aat att aca aga aat 727
Leu Pro Val Ser Arg Ile Thr Asp Leu Leu Pro Asn Ile Thr Arg Asn

70

tgagacacct gtgaccattc gaagtctgaa atgcgggggg ggagtttcat ttttaagtga 1867
 agacccaaaag ccctttaaaa ataatagttt tttatcattt tatagtaatc agcattttct 1927
 cttttactaa tatactcatt ccttttgaaa aaaaaaaaaa aa 1969

5 <210> 60
 <211> 1132
 <212> DNA
 <213> Homo sapiens

10 <220>
 <221> CDS
 <222> 77..937

<220>
 15 <221> sig_peptide
 <222> 77..127
 <223> Von Heijne matrix
 score 3.74817238048175
 seq RIVSAALLAFVQT/HL

20 <400> 60
 gttggtgggg ctgggggatg agagctgcac cgcgcgggac aagtcgccgg cgccccgacg 60
 gagcagaaga gagagc atg gag ctg gag agg atc gtc agt gca gcc ctc ctt 112
 Met Glu Leu Glu Arg Ile Val Ser Ala Ala Leu Leu
 25 -15 -10
 gcc ttt gtc cag aca cac ctc ccg gag gcc gac ctc agt ggc ttg gat 160
 Ala Phe Val Gln Thr His Leu Pro Glu Ala Asp Leu Ser Gly Leu Asp
 -5 1 5 10
 gag gtc atc ttc tcc tat gtg ctt ggg gtc ctg gag gac ctg ggc ccc 208
 30 Glu Val Ile Phe Ser Tyr Val Leu Gly Val Leu Glu Asp Leu Gly Pro
 15 20 25
 tcg ggc cca tca gag gag aac ttc gat atg gag gct ttc act gag atg 256
 Ser Gly Pro Ser Glu Glu Asn Phe Asp Met Glu Ala Phe Thr Glu Met
 30 35 40
 atg gag gcc tat gtg cct ggc ttc gcc cac atc ccc agg ggc aca ata 304
 Met Glu Ala Tyr Val Pro Gly Phe Ala His Ile Pro Arg Gly Thr Ile
 45 50 55
 ggg gac atg atg cag aag ctc tca ggg cag ctg agc gat gcc agg aac 352
 Gly Asp Met Met Gln Lys Leu Ser Gly Gln Leu Ser Asp Ala Arg Asn
 40 60 65 70 75
 aaa gag aac ctg caa ccg cag agc tct ggt gtc caa ggt cag gtg ccc 400
 Lys Glu Asn Leu Gln Pro Gln Ser Ser Gly Val Gln Gly Gln Val Pro
 80 85 90
 atc tcc cca gag ccc ctg cag cgg ccc gaa atg ctc aaa gaa gag act 448
 45 Ile Ser Pro Glu Pro Leu Gln Arg Pro Glu Met Leu Lys Glu Glu Thr
 95 100 105
 agg tct tcg gct gct gct gct gca gac acc caa gat gag gca act ggc 496
 Arg Ser Ser Ala Ala Ala Ala Ala Asp Thr Gln Asp Glu Ala Thr Gly
 110 115 120
 50 gct gag gag gag ctt ctg cca ggg gtg gat gta ctc ctg gag gtg ttc 544
 Ala Glu Glu Glu Leu Leu Pro Gly Val Asp Val Leu Leu Glu Val Phe
 125 130 135
 cct acc tgt tcg gtg gag cag gcc cag tgg gtg ctg gcc aaa gct cgg 592
 Pro Thr Cys Ser Val Glu Gln Ala Gln Trp Val Leu Ala Lys Ala Arg
 140 145 150 155
 ggg gac ttg gaa gaa gct gtg cag atg ctg gta gag gga aag gaa gag 640
 Gly Asp Leu Glu Glu Ala Val Gln Met Leu Val Glu Gly Lys Glu Glu
 160 165 170
 ggg cct gca gcc tgg gag ggc ccc aac cag gac ctg ccc aga cgc ctc 688
 60 Gly Pro Ala Ala Trp Glu Gly Pro Asn Gln Asp Leu Pro Arg Arg Leu
 175 180 185
 aga ggc ccc caa aag gat gag ctg aag tcc ttc atc ctg cag aag tac 736
 Arg Gly Pro Gln Lys Asp Glu Leu Lys Ser Phe Ile Leu Gln Lys Tyr


```

      190      195      200
atg atg gtg gat agc gca gag gat cag aag att cac cgg ccc atg gct      784
Met Met Val Asp Ser Ala Glu Asp Gln Lys Ile His Arg Pro Met Ala
      205      210      215
5  ccc aag gag gcc ccc aag aag ctg atc cga tac atc gac aac cag gta      832
Pro Lys Glu Ala Pro Lys Lys Leu Ile Arg Tyr Ile Asp Asn Gln Val
      220      225      230      235
gtg agc acc aaa ggg gag cga ttc aaa gat gtg cgg aac cct gag gcc      880
Val Ser Thr Lys Gly Glu Arg Phe Lys Asp Val Arg Asn Pro Glu Ala
10      240      245      250
gag gag atg aag gcc aca tac atc aac ctc aag cca gcc aga aag tac      928
Glu Glu Met Lys Ala Thr Tyr Ile Asn Leu Lys Pro Ala Arg Lys Tyr
      255      260      265
cgc ttc cat tgaggcactc gccggactct gcccgagcct tctaggctca      977
15 Arg Phe His
      270
gatcccagag ggatgcagga gccctataacc cctacacagg ggccccctaa ctctgtccc      1037
ccttctctac tcctttgctc catagtgtta acctactctc ggagctgcct ccatgggcac      1097
agtaaagggtg gcccaaggaa aaaaaaaaaa aaaaa      1132
20
<210> 61
<211> 631
<212> DNA
<213> Homo sapiens
25
<220>
<221> CDS
<222> 9..503
30
<220>
<221> sig_peptide
<222> 9..113
<223> Von Heijne matrix
score 10.2506494380376
35 seq LLPLVLLPPLAAA/AA

<400> 61
tgccaggg atg atg cgc tgc tgc cgc cgc cgc tgc tgc tgc cgg caa cca      50
Met Met Arg Cys Cys Arg Arg Arg Cys Cys Cys Arg Gln Pro
40      -35      -30      -25
ccc cat gcc ctg agg ccg ttg ctg ttg ctg ccc ctc gtc ctt tta cct      98
Pro His Ala Leu Arg Pro Leu Leu Leu Leu Pro Leu Val Leu Leu Pro
      -20      -15      -10
ccc ctg gca gca gct gca gcg ggc cca aac cga tgt gac acc ata tac      146
45 Pro Leu Ala Ala Ala Ala Gly Pro Asn Arg Cys Asp Thr Ile Tyr
      -5      1      5      10
cag ggc ttc gcc gag tgt ctc atc cgc ttg ggg gac agc atg ggc cgc      194
Gln Gly Phe Ala Glu Cys Leu Ile Arg Leu Gly Asp Ser Met Gly Arg
      15      20      25
50 gga ggc gag ctg gag acc atc tgc agg tct tgg aat tac ttc cat gcc      242
Gly Gly Glu Leu Glu Thr Ile Cys Arg Ser Trp Asn Tyr Phe His Ala
      30      35      40
tgt gcc tct cag gtc ctg tca ggc tgt ccg gag gag gca gct gca gtg      290
Cys Ala Ser Gln Val Leu Ser Gly Cys Pro Glu Glu Ala Ala Val
55      45      50      55
tgg gaa tca cta cag caa gaa gct cgc cag gcc ccc cgt ccg aat aac      338
Trp Glu Ser Leu Gln Gln Glu Ala Arg Gln Ala Pro Arg Pro Asn Asn
      60      65      70      75
ttg cac act ctg tgc ggt gcc ccg gtg cat gtt cgg gag cgc ggc aca      386
60 Leu His Thr Leu Cys Gly Ala Pro Val His Val Arg Glu Arg Gly Thr
      80      85      90
ggc tcc gaa acc aac cag gag acg ctg cgg gct aca gcg cct gca ctc      434
Gly Ser Glu Thr Asn Gln Glu Thr Leu Arg Ala Thr Ala Pro Ala Leu

```

```

      95              100              105
ccc atg gcc cct gcg ccc cca ctg ctg gcg gct gct ctg gct ctg gcc      482
Pro Met Ala Pro Ala Pro Pro Leu Leu Ala Ala Ala Leu Ala Leu Ala
      110              115              120
5  tac ctc ctg agg cct ctg gcc tagcttggtg ggttggttag cagcgcccg      533
   Tyr Leu Leu Arg Pro Leu Ala
      125              130
acctccagcc ctgctctggc ggtggttgctc caggctctgc agagcgcagc agggcttttc      593
attaaaggta tttatatttg caaaaaaaaaa aaaaaaaaaa      631
10
<210> 62
<211> 722
<212> DNA
<213> Homo sapiens
15
<220>
<221> CDS
<222> 21..464
20
<220>
<221> sig_peptide
<222> 21..95
<223> Von Heijne matrix.
      score 5.38058532480537
25      seq AVTSLLSPTPATA/LA

<400> 62
ggaagtgaagt gatcgaaagc atg gcg tcg gtg gtg ttg gcg ctg agg acc cgg      53
                        Met Ala Ser Val Val Leu Ala Leu Arg Thr Arg
30      -25              -20              -15
aca gcc gtt aca tcc ttg cta agc ccc act ccg gct aca gct ctt gct      101
Thr Ala Val Thr Ser Leu Leu Ser Pro Thr Pro Ala Thr Ala Leu Ala
      -10              -5              1
gtc aga tac gca tcc aag aag tcg ggt ggt agc tcc aaa aac ctc ggt      149
Val Arg Tyr Ala Ser Lys Lys Ser Gly Gly Ser Ser Lys Asn Leu Gly
      5              10              15
gga aag tca tca ggc aga cgc caa ggc att aag aaa atg gaa ggt cac      197
Gly Lys Ser Ser Gly Arg Arg Gln Gly Ile Lys Lys Met Glu Gly His
      20              25              30
40  tat gtt cat gct ggg aac atc att gca aca cag cgc cat ttc cgc tgg      245
Tyr Val His Ala Gly Asn Ile Ile Ala Thr Gln Arg His Phe Arg Trp
      35              40              45              50
cac cca ggt gcc cat gtg ggt gtt ggg aag aat aaa tgt ctg tat gcc      293
His Pro Gly Ala His Val Gly Val Gly Lys Asn Lys Cys Leu Tyr Ala
45      55              60              65
ctg gaa gag ggg ata gtc cgc tac act aag gag gtc tac gtg cct cat      341
Leu Glu Glu Gly Ile Val Arg Tyr Thr Lys Glu Val Tyr Val Pro His
      70              75              80
ccc aga aac acg gag gct gtg gat ctg atc acc agg ctg ccc aag ggt      389
Pro Arg Asn Thr Glu Ala Val Asp Leu Ile Thr Arg Leu Pro Lys Gly
      85              90              95
gct gtg ctc tac aag act ttt gtc cac gtg gtt cct gcc aag cct gag      437
Ala Val Leu Tyr Lys Thr Phe Val His Val Val Pro Ala Lys Pro Glu
      100              105              110
55  ggc acc ttc aaa ctg gta gct atg ctt tgatgtcctg ttgaggccat      484
   Gly Thr Phe Lys Leu Val Ala Met Leu
      115              120
cggacagaga ctggagccca ggtgacagga gatggtgata ccagaagtca agggttgggg      544
tggcgacacg gcctcccagag gaagagggtct gcttgatggt gactctgcag gagactctga      604
60  agtgactgct gggaaaccct ttgggagacc tgacctgggg ccaaaaataa agtgagccag      664
   cgtcatgaac gcatgctatt tagggacaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa      722

<210> 63

```

<211> 1442
 <212> DNA
 <213> Homo sapiens

5 <220>
 <221> CDS
 <222> 178..1050

<220>
 10 <221> sig_peptide
 <222> 178..279
 <223> Von Heijne matrix
 score 10.0571391689271
 seq FLCLLSALLLTEG/KK

15 <400> 63
 agtgcattgc tggagcgcagg agaagctcac gaatcagctg caggtctctg ttttgaaaaa 60
 gcagagatac agaggcagag gaaaagggca ctcctatgtg acctgttctt agagcaagac 120
 aatcaccatc tgaattccag aagccctgtt catgggtggg gatattttct cgactgc 177
 20 atg gaa tca gaa aga agc aaa agg atg gga aat gcc tgc att ccc ctg 225
 Met Glu Ser Glu Arg Ser Lys Arg Met Gly Asn Ala Cys Ile Pro Leu
 -30 -25 -20
 aaa aga att gct tat ttc cta tgt ctc tta tct gcg ctt ttg ctg act 273
 Lys Arg Ile Ala Tyr Phe Leu Cys Leu Leu Ser Ala Leu Leu Leu Thr
 -15 -10 -5
 25 gag ggg aag aaa cca gcg aag cca aaa tgc cct gcc gtg tgt act tgt 321
 Glu Gly Lys Lys Pro Ala Lys Pro Lys Cys Pro Ala Val Cys Thr Cys
 1 5 10
 acc aaa gat aat gct tta tgt gag aat gcc aga tcc att cca cgc acc 369
 30 Thr Lys Asp Asn Ala Leu Cys Glu Asn Ala Arg Ser Ile Pro Arg Thr
 15 20 25 30
 gtt cct cct gat gtt atc tca tta tcc ttt gtg aga tct ggt ttt act 417
 Val Pro Pro Asp Val Ile Ser Leu Ser Phe Val Arg Ser Gly Phe Thr
 35 40 45
 35 gaa atc tca gaa ggg agt ttt tta ttc acg cca tcg ctg cag ctc ttg 465
 Glu Ile Ser Glu Gly Ser Phe Leu Phe Thr Pro Ser Leu Gln Leu Leu
 50 55 60
 tta ttc aca tcg aac tcc ttt gat gtg atc agt gat gat gct ttt att 513
 Leu Phe Thr Ser Asn Ser Phe Asp Val Ile Ser Asp Asp Ala Phe Ile
 65 70 75
 40 ggt ctt cca cat cta gag tat tta ttc ata gaa aac aac aac atc aag 561
 Gly Leu Pro His Leu Glu Tyr Leu Phe Ile Glu Asn Asn Asn Ile Lys
 80 85 90
 tca att tca aga cat act ttc cgg gga cta aag tca tta att cac ttg 609
 45 Ser Ile Ser Arg His Thr Phe Arg Gly Leu Lys Ser Leu Ile His Leu
 95 100 105 110
 agc ctt gca aac aac aat ctc cag aca ctc cca aaa gat att ttc aaa 657
 Ser Leu Ala Asn Asn Asn Leu Gln Thr Leu Pro Lys Asp Ile Phe Lys
 115 120 125
 50 ggc ctg gat tct tta aca aat gtg gac ctg agg ggt aat tca ttt aat 705
 Gly Leu Asp Ser Leu Thr Asn Val Asp Leu Arg Gly Asn Ser Phe Asn
 130 135 140
 tgt gac tgt aaa ctg aaa tgg cta gtg gaa tgg ctt ggc cac acc aat 753
 Cys Asp Cys Lys Leu Lys Trp Leu Val Glu Trp Leu Gly His Thr Asn
 145 150 155
 55 gca act gtt gaa gac atc tac tgc gaa ggc ccc cca gaa tac aag aag 801
 Ala Thr Val Glu Asp Ile Tyr Cys Glu Gly Pro Pro Glu Tyr Lys Lys
 160 165 170
 cgc aaa atc aat agt ctc tcc tcg aag gat ttc gat tgc atc att aca 849
 60 Arg Lys Ile Asn Ser Leu Ser Ser Lys Asp Phe Asp Cys Ile Ile Thr
 175 180 185 190
 gaa ttt gca aag tct caa gac ctg cct tat caa tca ttg tcc ata gac 897
 Glu Phe Ala Lys Ser Gln Asp Leu Pro Tyr Gln Ser Leu Ser Ile Asp

					195					200						205		
	act	ttt	tct	tat	ttg	aat	gat	gag	tat	gta	gtc	atc	gct	cag	cct	ttt	945	
	Thr	Phe	Ser	Tyr	Leu	Asn	Asp	Glu	Tyr	Val	Val	Ile	Ala	Gln	Pro	Phe		
				210					215					220				
5	act	gga	aaa	tgc	att	ttc	ctt	gaa	tgg	gac	cat	gtg	gaa	aag	acc	ttc	993	
	Thr	Gly	Lys	Cys	Ile	Phe	Leu	Glu	Trp	Asp	His	Val	Glu	Lys	Thr	Phe		
			225					230						235				
	cgg	aat	tat	gac	aac	att	aca	gtt	tta	agg	gaa	ata	cac	aga	ttt	aca	1041	
	Arg	Asn	Tyr	Asp	Asn	Ile	Thr	Val	Leu	Arg	Glu	Ile	His	Arg	Phe	Thr		
10		240				245						250						
	aac	atg	tca	tagttgactt	aagcgc	catga	gacaccaa	at	tctgtggctg								1090	
	Asn	Met	Ser															
	255																	
	ccatcagaaa	ttttctacag	tacatgaccc	ggatgaactc	aatgcatgat	gactcttctt											1150	
15	atcacacttg	caaataaatg	cctttcaaac	attgagactg	ctagaacca	gcactaccag											1210	
	tatctccatc	cttaactgtc	cagtccagtg	atgtgggaag	ttacctttta	taagacaaaa											1270	
	tttaattgtg	taactgttct	ttgcagtgaa	gatgtgtaaa	taagcgttta	atggtatctg											1330	
	ttactccaaa	aagaaatatt	aatatgtact	tttccattta	tttattcatg	tgtacagaaa											1390	
	caactgccaa	ataaaatggt	tacattttct	ttcagaaaaa	aaaaaaaaaa	aa											1442	
20																		
	<210>	64																
	<211>	795																
	<212>	DNA																
	<213>	Homo sapiens																
25																		
	<220>																	
	<221>	CDS																
	<222>	32..274																
30	<220>																	
	<221>	sig_peptide																
	<222>	32..178																
	<223>	Von Heijne matrix																
	score	4.30837886795471																
35		seq	LMVELLKVFVVEA/AV															
	<400>	64																
	gttccggtgg	gcgcgcgttg	aggctgcggt	c	atg	gag	gga	gca	gga	gct	gga						52	
					Met	Glu	Gly	Ala	Gly	Ala	Gly							
40																		
	tcc	ggc	ttc	cgg	aag	gag	ctg	gtg	agc	agg	ctg	ctg	cac	ctg	cac	ttc	100	
	Ser	Gly	Phe	Arg	Lys	Glu	Leu	Val	Ser	Arg	Leu	Leu	His	Leu	His	Phe		
		-40						-35					-30					
	aag	gat	gac	aag	acc	aaa	gtg	agc	ggg	gac	gcg	ctg	cag	ctc	atg	gtg	148	
45	Lys	Asp	Asp	Lys	Thr	Lys	Val	Ser	Gly	Asp	Ala	Leu	Gln	Leu	Met	Val		
		-25					-20					-15						
	gag	ttg	ctg	aag	gtc	ttc	gtt	gtg	gaa	gca	gca	gtc	cgc	ggc	gtg	cgg	196	
	Glu	Leu	Leu	Lys	Val	Phe	Val	Val	Glu	Ala	Ala	Val	Arg	Gly	Val	Arg		
		-10				-5					1				5			
50	cag	gcc	cag	gca	gaa	gac	gcg	ctc	cgt	gtg	gac	gtg	gac	cag	ctg	gag	244	
	Gln	Ala	Gln	Ala	Glu	Asp	Ala	Leu	Arg	Val	Asp	Val	Asp	Gln	Leu	Glu		
				10					15					20				
	aag	gtg	ctt	ccg	cag	ctg	ctc	ctg	gac	ttc	tagggatctc	agccgtggct					294	
	Lys	Val	Leu	Pro	Gln	Leu	Leu	Leu	Asp	Phe								
55		25						30										
	gaggccaccc	ccagaggagc	ccctgggtcca	cagaagcagg	ccttgtgttt	ccagcggcct											354	
	ctgataagag	gcaggggaagg	acctgaagga	tttggagttg	attcaaacia	gatctctggg											414	
	agtctccagc	ctgtgcagaa	ggggcaggac	tgcagtgcac	tgcgggcctt	ggagtgtcca											474	
	gtggggacac	tggtgtggga	aggggcagca	cctggggagt	ccctgcctct	cctccctggg											534	
60	acaatagtgt	gcatgccacc	cggggtccta	caggcagggtg	ctgggaaagg	cctggccagc											594	
	aggtagcctg	tgtgtttgac	aaacagcagc	tggcagcgct	gcctcctgcc	cacattcctg											654	
	ccacccgaca	tcaaagctgg	cgtgtgacct	ttccagccat	gcgatattcc	ccttgggaaga											714	
	tqcttcccca	qqctataaat	ttgttctcac	aaagcaacat	caataaatca	aaactgtctc											774	

tctcaaaaaa aaaaaaaaaa a 795

<210> 65
 <211> 1236
 5 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 10 <222> 222..920

<220>
 <221> sig_peptide
 <222> 222..311
 15 <223> Von Heijne matrix
 score 4.35083245061594
 seq VAHALSLPAESYG/ND

<400> 65
 20 accgaaaatt actgacgagt caatcacctc agatctctca agcagtccag cctacgcaac 60
 agtactccac ctctgcgcct gtgcggggag ggtaaggcgg ggccagcaac ttcctcagct 120
 ggagggagag cgcacggtgg agccgccagt tgagaaggac tctgatccgg ctcagctttc 180
 caatcagctg cggaaggagc cacgctttcg ggggttgcaa g atg gcg gcc acc agt 236
 Met Ala Ala Thr Ser
 -30

25 gga act gat gag ccg gtt tcc ggg gag ttg gtg tct gtg gca cat gcg 284
 Gly Thr Asp Glu Pro Val Ser Gly Glu Leu Val Ser Val Ala His Ala
 -25 -20 -15 -10
 ctt tct ctc cca gca gag tcg tat ggc aac gat cct gac att gag atg 332
 30 Leu Ser Leu Pro Ala Glu Ser Tyr Gly Asn Asp Pro Asp Ile Glu Met
 -5 1 5
 gct tgg gcc atg aga gca atg cag cat gct gaa gtc tat tac aag ctg 380
 Ala Trp Ala Met Arg Ala Met Gln His Ala Glu Val Tyr Tyr Lys Leu
 10 15 20
 35 att tca tca gtt gac cca cag ttc ctc aaa ctc acc aaa gta gat gac 428
 Ile Ser Ser Val Asp Pro Gln Phe Leu Lys Leu Thr Lys Val Asp Asp
 25 30 35
 caa att tac tct gag ttc cgg aaa aat ttt gag acc ctt agg ata gat 476
 Gln Ile Tyr Ser Glu Phe Arg Lys Asn Phe Glu Thr Leu Arg Ile Asp
 40 40 45 50 55
 gtg ttg gac cca gaa gaa ctc aag tca gaa tca gcc aaa gag aag tgg 524
 Val Leu Asp Pro Glu Glu Leu Lys Ser Glu Ser Ala Lys Glu Lys Trp
 60 65 70
 45 agg cca ttc tgc ttg aag ttt aat ggg att gtt gaa gac ttc aac tat 572
 Arg Pro Phe Cys Leu Lys Phe Asn Gly Ile Val Glu Asp Phe Asn Tyr
 75 80 85
 ggt act ttg ctg cga cta gat tgt tct cag ggc tac act gag gaa aac 620
 Gly Thr Leu Leu Arg Leu Asp Cys Ser Gln Gly Tyr Thr Glu Glu Asn
 90 95 100
 50 acc atc ttt gcc ccc agg ata caa ttc ttt gcc att gaa att gct cgg 668
 Thr Ile Phe Ala Pro Arg Ile Gln Phe Phe Ala Ile Glu Ile Ala Arg
 105 110 115
 aac cgg gaa ggc tat aac aaa gct gtt tat atc agt gtt cag gac aaa 716
 Asn Arg Glu Gly Tyr Asn Lys Ala Val Tyr Ile Ser Val Gln Asp Lys
 55 120 125 130 135
 gaa gga gag aaa gga gtc aac aat gga gga gaa aaa aga gct gac agt 764
 Glu Gly Glu Lys Gly Val Asn Asn Gly Gly Glu Lys Arg Ala Asp Ser
 140 145 150
 60 gga gaa gaa gag aac acc aag aat gga gga gag aaa gga gct gat agt 812
 Gly Glu Glu Glu Asn Thr Lys Asn Gly Gly Glu Lys Gly Ala Asp Ser
 155 160 165
 gga gaa gaa aaa gag gaa gga atc aac aga gaa gac aaa act gac aaa 860
 Gly Glu Glu Lys Glu Glu Gly Ile Asn Arg Glu Asp Lys Thr Asp Lys

```

      170      175      180
gga gga gaa aaa ggg aaa gaa gct gac aaa gaa atc aac aaa agt ggt      908
Gly Gly Glu Lys Gly Lys Glu Ala Asp Lys Glu Ile Asn Lys Ser Gly
      185      190      195
5  gaa aaa gct atg taaggatatac agggaacagc actctagaag ctatgactca      960
   Glu Lys Ala Met
      200
attgagacta caagtaccac ggtgctactt gcacagaccc ctttgggttaa atgtaaattc      1020
ttgtacaatt gaaggatacg cagaaggaca tctttctagt ctaacagtca ggagctgctc      1080
10  tgggtcattcc cttgtatgaa ctgggtctaaa gactgttagt ggggtgtag ttgatttttc      1140
   ctggtatact gtttcttggc tgacactact ggtcaagtaa gaaatttgta aataaatttc      1200
   ttttggttct tattaaaaca aaaaaaaaaa aaaaaa      1236

<210> 66
15 <211> 881
   <212> DNA
   <213> Homo sapiens

<220>
20 <221> CDS
   <222> 101..355

<220>
<221> sig_peptide
25 <222> 101..160
   <223> Von Heijne matrix
       score 9.32665652007071
       seq LFLCYLLLFTCSG/VE

30 <400> 66
   ttactcgctg ctgtgcccac ctatcagcag gctccgggct gaagattgct tctcttctct      60
   cctccaaggt ctagtgcagg agcccgcgcg cggcgccacc atg cgg cag aag gcg      115
                                     Met Arg Gln Lys Ala
                                     -20
35 gta tcg ctt ttc ttg tgc tac ctg ctg ctc ttc act tgc agt ggg gtg      163
   Val Ser Leu Phe Leu Cys Tyr Leu Leu Leu Phe Thr Cys Ser Gly Val
   -15      -10      -5      1
   gag gca ggt aag aaa aag tgc tcg gag agc tcg gac agc ggc tcc ggg      211
   Glu Ala Gly Lys Lys Lys Cys Ser Glu Ser Ser Asp Ser Gly Ser Gly
   40      5      10      15
   ttc tgg aag gcc ctg acc ttc atg gcc gtc gga gga gga ctc gca gtc      259
   Phe Trp Lys Ala Leu Thr Phe Met Ala Val Gly Gly Gly Leu Ala Val
   20      25      30
   gcc ggg ctg ccc gcg ctg ggc ttc acc ggc gcc ggc atc gcg gcc aac      307
45  Ala Gly Leu Pro Ala Leu Gly Phe Thr Gly Ala Gly Ile Ala Ala Asn
   35      40      45
   tcg gtg gct gcc tcg ctg atg agc tgg tct gcg atc ctg aat ggg ggc      355
   Ser Val Ala Ala Ser Leu Met Ser Trp Ser Ala Ile Leu Asn Gly Gly
   50      55      60      65
50  tagtggccac gctgcagagc ctcggggctg gtggcagcag cgtcgtcata ggtaatatg      415
   gtgccctgat gggctacgcc acccacaagt atctcgatag tgaggaggat gaggagtagc      475
   cagcagctcc cagaacctct tcttccttct tggcctaact cttccagtta ggatctagaa      535
   ctttgccctt tttttttttt tttttttttt ttgagatggg ttctcactat attgtccagg      595
   ctagagtgca gkggctattc acagatgcga acatagtaca ctgcagcctc caactcctag      655
55  cctcaagtga tcctcctgtc tcaacctccc aagtaggatt acaagcatgc gccgacgatg      715
   cccaraatcc araactttgt ctatcactct ccccaacaac ctagatgtga aaacagaata      775
   aacttcaccc agaaaaaaaa aaammacaar aaaaaaaaaa aaaaaaaaaa aaaaaaaaaaam      835
   aaaaaaaaaa rrrraaaaaa aaaaaaaaga aaaaaaaaaa aaaaaa      881

60 <210> 67
   <211> 524
   <212> DNA
   <213> Homo sapiens

```

<220>
 <221> CDS
 <222> 173..487
 5
 <220>
 <221> sig_peptide
 <222> 173..301
 <223> Von Heijne matrix
 10 score 4.27484469223909
 seq AGSLVATLQSVGA/AG

 <400> 67
 15 agggcagagt aggcgcgtcc ctactggatg gagggggaag taacacccca agaacgctgt 60
 catttcctgg gccaaagttgg gacccggacg gcctcaccat gatgaaacgg gcagctgctg 120
 ctgcagtggg aggagccctg gcagtggggg ctgtgccctg gtgctcagtg cc atg ggc 178
 Met Gly
 ttc act ggg gca gga atc gcc gcg tcc tcc ata gca gcc aag atg atg 226
 Phe Thr Gly Ala Gly Ile Ala Ala Ser Ser Ile Ala Ala Lys Met Met
 20 -40 -35 -30
 tcc gca gca gcc att gcc aac ggg ggt ggt gtt tct gcg ggg agc ctg 274
 Ser Ala Ala Ala Ile Ala Asn Gly Gly Gly Val Ser Ala Gly Ser Leu
 -25 -20 -15 -10
 gtg gct act ctg cag tcc gtg ggg gca gct gga ctc tcc aca tca tcc 322
 25 Val Ala Thr Leu Gln Ser Val Gly Ala Ala Gly Leu Ser Thr Ser Ser
 -5 1 5
 aac atc ctc ctg gcc tct gtt ggg tca gtg ttg ggg gcc tgc ttg ggg 370
 Asn Ile Leu Leu Ala Ser Val Gly Ser Val Leu Gly Ala Cys Leu Gly
 10 15 20
 30 aat tca cct tct tct tct ctc cca gct gaa ccc gag gct aaa gaa gat 418
 Asn Ser Pro Ser Ser Ser Leu Pro Ala Glu Pro Glu Ala Lys Glu Asp
 25 30 35
 gag gca aga gaa aat gta ccc caa ggt gaa cct cca aaa ccc cca ctc 466
 Glu Ala Arg Glu Asn Val Pro Gln Gly Glu Pro Pro Lys Pro Pro Leu
 35 40 45 50 55
 aag tca gag aaa cat gag gaa taaaggtcac atgcagatgc aaaaaaaaaa 517
 Lys Ser Glu Lys His Glu Glu
 60
 aaaaaaa 524
 40
 <210> 68
 <211> 1472
 <212> DNA
 <213> Homo sapiens
 45
 <220>
 <221> CDS
 <222> 210..1082
 50
 <220>
 <221> sig_peptide
 <222> 210..311
 <223> Von Heijne matrix
 score 10.0571391689271
 55 seq FLCLLSALLLTEG/KK

 <400> 68
 60 acagtacctc acaggtctct tcccccgagc agtgcattgc tggagcgagg agaagctcac 60
 gaatcagctg caggtctctg ttttgaaaaa gcagagatac agaggcagag gaaaaggggtg 120
 gactcctatg tgacctgttc ttagagcaag acaatcacca tctgaattcc agaagccctg 180
 ttcatggttg gggatatatt ctcgactgc atg gaa tca gaa aga agc aaa agg 233
 Met Glu Ser Glu Arg Ser Lys Arg
 -30

	atg gga aat gcc tgc att ccc ctg aaa aga att gct tat ttc cta tgt	281
	Met Gly Asn Ala Cys Ile Pro Leu Lys Arg Ile Ala Tyr Phe Leu Cys	
	-25 -20 -15	
5	ctc tta tct gcg ctt ttg ctg act gag ggg aag aaa cca gcg aag cca	329
	Leu Leu Ser Ala Leu Leu Leu Thr Glu Gly Lys Lys Pro Ala Lys Pro	
	-10 -5 1 5	
	aaa tgc cct gcc gtg tgt act tgt acc aaa gat aat gct tta tgt gag	377
	Lys Cys Pro Ala Val Cys Thr Cys Thr Lys Asp Asn Ala Leu Cys Glu	
	10 10 15 20	
10	aat gcc aga tcc att cca cgc acc gtt cct cct gat gtt atc tca tta	425
	Asn Ala Arg Ser Ile Pro Arg Thr Val Pro Pro Asp Val Ile Ser Leu	
	25 30 35	
	tcc ttt gtg aga tct gtt ttt act gaa atc tca gaa ggg agt ttt tta	473
	Ser Phe Val Arg Ser Val Phe Thr Glu Ile Ser Glu Gly Ser Phe Leu	
15	40 45 50	
	ttc acg cca tcg ctg cag ctc ttg tta ttc aca tcg aac tcc ttt gat	521
	Phe Thr Pro Ser Leu Gln Leu Leu Leu Phe Thr Ser Asn Ser Phe Asp	
	55 60 65 70	
20	gtg atc agt gat gat gct ttt att ggt ctt cca cat cta gag tat tta	569
	Val Ile Ser Asp Asp Ala Phe Ile Gly Leu Pro His Leu Glu Tyr Leu	
	75 80 85	
	ttc ata gaa aac aac aac atc aag tca att tca aga cat act ttc cgg	617
	Phe Ile Glu Asn Asn Asn Ile Lys Ser Ile Ser Arg His Thr Phe Arg	
	90 95 100	
25	gga cta aag tca tta att cac ttg agc ctt gca aac aac aat ctc cag	665
	Gly Leu Lys Ser Leu Ile His Leu Ser Leu Ala Asn Asn Asn Leu Gln	
	105 110 115	
	aca ctc cca aaa gat att ttc aaa ggc ctg gat tct tta aca aat gtg	713
	Thr Leu Pro Lys Asp Ile Phe Lys Gly Leu Asp Ser Leu Thr Asn Val	
30	120 125 130	
	gac ctg agg ggt aat tca ttt aat tgt gac tgt aaa ctg aaa tgg cta	761
	Asp Leu Arg Gly Asn Ser Phe Asn Cys Asp Cys Lys Leu Lys Trp Leu	
	135 140 145 150	
35	gtg gaa tgg ctt ggc cac acc aat gca act gtt gaa gac atc tac tgc	809
	Val Glu Trp Leu Gly His Thr Asn Ala Thr Val Glu Asp Ile Tyr Cys	
	155 160 165	
	gaa ggc ccc cca gaa tac aag aag cgc aaa atc aat agt ctc tcc tcg	857
	Glu Gly Pro Pro Glu Tyr Lys Lys Arg Lys Ile Asn Ser Leu Ser Ser	
	170 175 180	
40	aag gat ttc gat tgc atc att aca gaa ttt gca aag tct caa gac ctg	905
	Lys Asp Phe Asp Cys Ile Ile Thr Glu Phe Ala Lys Ser Gln Asp Leu	
	185 190 195	
	cct tat caa tca ttg tcc ata gac act ttt tct tat ttg aat gat gag	953
	Pro Tyr Gln Ser Leu Ser Ile Asp Thr Phe Ser Tyr Leu Asn Asp Glu	
45	200 205 210	
	tat gta gtc atc gct cag cct ttt act gga aaa tgc att ttc ctt gaa	1001
	Tyr Val Val Ile Ala Gln Pro Phe Thr Gly Lys Cys Ile Phe Leu Glu	
	215 220 225 230	
50	tgg gac cat gtg gaa aag acc ttc cgg aat tat gac aac att aca gtt	1049
	Trp Asp His Val Glu Lys Thr Phe Arg Asn Tyr Asp Asn Ile Thr Val	
	235 240 245	
	tta agg gaa ata cac aga ttt aca aac atg tca tagttgactt aagcgcata	1102
	Leu Arg Glu Ile His Arg Phe Thr Asn Met Ser	
	250 255	
55	gacaccaaatt tctgtggctg ccatacagaaa ttttctacag tacatgaccc ggatgaactc	1162
	aatgcatgat gactcttctt atcacacttg caaatgaatg cctttcaaac attgagactg	1222
	ctagaaccaa gcactaccag tatctccatc cttaactgtc cagtccagtg atgtgggaag	1282
	ttacctttta taagacaaaa tttaattgtg taactgttct ttgcagtga gatgtgtaaa	1342
	taagcgttta atggtatctg ttactccaaa aagaaatatt aatatgtact tttccattta	1402
60	tttattcatg tgtacagaaa caactgccaa ataaaatgtt tacattttct tacaacaaaa	1462
	aaaaaaaaa	1472

<210> 69

<211> 1737
 <212> DNA
 <213> Homo sapiens

5 <220>
 <221> CDS
 <222> 172..1449

<220>
 10 <221> sig_peptide
 <222> 172..255
 <223> Von Heijne matrix
 score 5.94825670923113
 seq XVLLEPFVHQVGG/HS

15 <400> 69
 aaacaatagg acggaacgc cgaggaaccc ggctgaggcg gcagagcatc ctggccagaa 60
 caagccaagg agccaagacg agagggacac acggacaaac aacagacaga agacgtactg 120
 gccgctggac tccgctgcct ccccatctc cccgccatct gcgcccggag g atg agc 177
 Met Ser
 20 cca gcc ttc agg gcc atg gat gtg gag ccc cgc gcc aaa ggc gtc ctt 225
 Pro Ala Phe Arg Ala Met Asp Val Glu Pro Arg Ala Lys Gly Val Leu
 -25 -20 -15
 ctg gag ccc ttt gtc cac cag gtc ggg ggg cac tca tgc gtg ctc cgc 273
 25 Leu Glu Pro Phe Val His Gln Val Gly Gly His Ser Cys Val Leu Arg
 -10 -5 1 5
 ttc aat gag aca acc ctg tgc aag ccc ctg gtc cca agg gaa cat cag 321
 Phe Asn Glu Thr Leu Cys Lys Pro Leu Val Pro Arg Glu His Gln
 10 15 20
 30 ttc tac gag acc ctc cct gct gag atg cgc aaa ttc act ccc cag tac 369
 Phe Tyr Glu Thr Leu Pro Ala Glu Met Arg Lys Phe Thr Pro Gln Tyr
 25 30 35
 aaa ggt gtg gta tct gtg cgc ttt gaa gaa gat gaa gac agg aac ttg 417
 Lys Gly Val Val Ser Val Arg Phe Glu Glu Asp Glu Asp Arg Asn Leu
 40 45 50
 35 tgt cta ata gca tat cca ttg aaa ggg gac cat gga att gtg gac att 465
 Cys Leu Ile Ala Tyr Pro Leu Lys Gly Asp His Gly Ile Val Asp Ile
 55 60 65 70
 gta gat aat tca gac tgt gaa cca aaa agt aag ctc cta agg tgg aca 513
 40 Val Asp Asn Ser Asp Cys Glu Pro Lys Ser Lys Leu Leu Arg Trp Thr
 75 80 85
 aca aac aaa aaa cat cat gtc tta gaa aca gaa aag acc cct aag gac 561
 Thr Asn Lys Lys His His Val Leu Glu Thr Glu Lys Thr Pro Lys Asp
 90 95 100
 45 tgg gtg cgt cag cac cgt aaa gag gag aaa atg aag agc cat aag tta 609
 Trp Val Arg Gln His Arg Lys Glu Glu Lys Met Lys Ser His Lys Leu
 105 110 115
 gaa gaa gaa ttt gag tgg cta aag aaa tct gaa gtc ttg tac tac act 657
 Glu Glu Glu Phe Glu Trp Leu Lys Lys Ser Glu Val Leu Tyr Tyr Thr
 120 125 130
 50 gta gag aag aag ggg aat ata agt tcc cag ctt aaa cac tat aac cct 705
 Val Glu Lys Lys Gly Asn Ile Ser Ser Gln Leu Lys His Tyr Asn Pro
 135 140 145 150
 tgg agc atg aaa tgt cac cag caa cag tta cag aga atg aag gag aat 753
 55 Trp Ser Met Lys Cys His Gln Gln Gln Leu Gln Arg Met Lys Glu Asn
 155 160 165
 gca aag cat cgg aac cag tac aaa ttt atc tta ctg gaa aac ctg act 801
 Ala Lys His Arg Asn Gln Tyr Lys Phe Ile Leu Leu Glu Asn Leu Thr
 170 175 180
 60 tcc cgc tat gag gtg cct tgt gtc ctt gac ctc aag atg ggc aca cga 849
 Ser Arg Tyr Glu Val Pro Cys Val Leu Asp Leu Lys Met Gly Thr Arg
 185 190 195
 caa cat ggt gat gat gct tca gag gag aag gca gcc aac cag atc cga 897

```

      Gln His Gly Asp Asp Ala Ser Glu Glu Lys Ala Ala Asn Gln Ile Arg
      200                205                210
      aaa tgt cag cag agc aca tct gca gtc att ggt gtg cgt gtg tgt ggc      945
      Lys Cys Gln Gln Ser Thr Ser Ala Val Ile Gly Val Arg Val Cys Gly
5    215                220                225                230
      atg cag gtg tac caa gca ggc agt ggg cag ctc atg ttc atg aac aag      993
      Met Gln Val Tyr Gln Ala Gly Ser Gly Gln Leu Met Phe Met Asn Lys
      235                240                245
      tac cat gga cgg aag cta tcg atg cag ggc ttc aag gag gca ctt ttc      1041
10   Tyr His Gly Arg Lys Leu Ser Met Gln Gly Phe Lys Glu Ala Leu Phe
      250                255                260
      cag ttc ttc cac aat ggg cgg tac ctg cgc cgt gaa ctc ctg ggc cct      1089
      Gln Phe Phe His Asn Gly Arg Tyr Leu Arg Arg Glu Leu Leu Gly Pro
      265                270                275
15   gtg ctc aag aag ctg act gag ctc aag gca gtg ttg gag cga cag gag      1137
      Val Leu Lys Lys Leu Thr Glu Leu Lys Ala Val Leu Glu Arg Gln Glu
      280                285                290
      tcc tac cgc ttc tac tca agc tcc ctg ctg gtc att tat gat ggc aag      1185
      Ser Tyr Arg Phe Tyr Ser Ser Ser Leu Leu Val Ile Tyr Asp Gly Lys
20   295                300                305                310
      gag cgg ccc gaa gtg gtc ctg gac tca gat gct gag gat ttg gag gac      1233
      Glu Arg Pro Glu Val Val Leu Asp Ser Asp Ala Glu Asp Leu Glu Asp
      315                320                325
      ctg tca gag gaa tca gct gat gag tct gct ggt gcc tat gcc tac aaa      1281
25   Leu Ser Glu Glu Ser Ala Asp Glu Ser Ala Gly Ala Tyr Ala Tyr Lys
      330                335                340
      ccc atc ggc gcc agc tct gta gat gtg cgc atg atc gac ttt gca cac      1329
      Pro Ile Gly Ala Ser Ser Val Asp Val Arg Met Ile Asp Phe Ala His
      345                350                355
30   acc acc tgc agg ctg tat ggc gag gac acc gtg gtg cat gag ggc cag      1377
      Thr Thr Cys Arg Leu Tyr Gly Glu Asp Thr Val Val His Glu Gly Gln
      360                365                370
      gat gct ggc tat atc ttc ggg ctc cag agc ctg ata gac att gtc aca      1425
      Asp Ala Gly Tyr Ile Phe Gly Leu Gln Ser Leu Ile Asp Ile Val Thr
35   375                380                385                390
      gag ata agt gag gag agt ggg gag tgagcttgct agctgctcca gtacttgaga      1479
      Glu Ile Ser Glu Glu Ser Gly Glu
      395
40   gcgactctgt gtcccaggca cagctgtgct gcgtcagggg ggaagccagt atggccaggt      1539
      ggtgggtcct gcagcctgga gctgatgtgc agtggcctct gtgagcccca gcctgagcca      1599
      gtcccagctg tgcttgaggt ctttatttat tttaactatt tcttcaacat tccacatttg      1659
      atgatgatac ctctttcttc cctgagtgtg tatgtttctaa tacaaatctt tttgtttatt      1719
      gaaaaaaaa aaaaaaaaaa      1737

45 <210> 70
    <211> 1637
    <212> DNA
    <213> Homo sapiens

50 <220>
    <221> CDS
    <222> 30..1427

    <220>
55 <221> sig_peptide
    <222> 30..77
    <223> Von Heijne matrix
        score 3.71064775937629
        seq YAAAAGVLAGVES/RQ

60 <400> 70
    ctaatcgaaa agtaaaggcg cgcgggaac atg ggg ctg tat gct gca gct gca      53
    Met Gly Leu Tyr Ala Ala Ala Ala

```

									-15					-10			
	ggc	gtg	ttg	gcc	ggc	gtg	gag	agc	cgc	cag	ggc	tct	atc	aag	ggg	ttg	101
	Gly	Val	Leu	Ala	Gly	Val	Glu	Ser	Arg	Gln	Gly	Ser	Ile	Lys	Gly	Leu	
				-5					1				5				
5	gtg	tac	tcc	agc	aac	ttc	cag	aac	gtg	aag	cag	ctg	tac	gcg	ctg	gtg	149
	Val	Tyr	Ser	Ser	Asn	Phe	Gln	Asn	Val	Lys	Gln	Leu	Tyr	Ala	Leu	Val	
	10						15					20					
	tgc	gaa	acg	cag	cgc	tac	tcc	gcc	gtg	ctg	gat	gct	gtg	atc	gcc	agc	197
	Cys	Glu	Thr	Gln	Arg	Tyr	Ser	Ala	Val	Leu	Asp	Ala	Val	Ile	Ala	Ser	
10	25					30					35				40		
	gcc	ggc	ctc	ctc	cgt	gcg	gag	aag	aag	ctg	cgg	ccg	cac	ctg	gcc	aag	245
	Ala	Gly	Leu	Leu	Arg	Ala	Glu	Lys	Lys	Leu	Arg	Pro	His	Leu	Ala	Lys	
					45					50					55		
	gtg	cta	gtg	tat	gag	ttg	ttg	ttg	gga	aag	ggc	ttt	cga	ggg	ggg	ggg	293
15	Val	Leu	Val	Tyr	Glu	Leu	Leu	Leu	Gly	Lys	Gly	Phe	Arg	Gly	Gly	Gly	
				60					65					70			
	ggc	cga	tgg	aag	gct	ctg	ttg	ggc	cgg	cac	cag	gcg	agg	ctc	aag	gct	341
	Gly	Arg	Trp	Lys	Ala	Leu	Leu	Gly	Arg	His	Gln	Ala	Arg	Leu	Lys	Ala	
				75				80					85				
20	gag	ttg	gct	cgg	ctc	aag	gtt	cat	cgg	ggg	gtg	agc	cgg	aat	gag	gac	389
	Glu	Leu	Ala	Arg	Leu	Lys	Val	His	Arg	Gly	Val	Ser	Arg	Asn	Glu	Asp	
		90					95					100					
	ctg	ttg	gaa	gtg	gga	tcc	agg	cct	ggg	cca	gcc	tcc	cag	ctg	cct	cga	437
	Leu	Leu	Glu	Val	Gly	Ser	Arg	Pro	Gly	Pro	Ala	Ser	Gln	Leu	Pro	Arg	
25	105					110					115					120	
	ttt	gtg	cgt	gtg	aac	act	ctc	aag	acc	tgc	tcc	gat	gat	gta	gtt	gat	485
	Phe	Val	Arg	Val	Asn	Thr	Leu	Lys	Thr	Cys	Ser	Asp	Asp	Val	Val	Asp	
					125					130					135		
	tat	ttc	aag	aga	caa	ggg	ttc	tcc	tat	cag	ggg	cgg	gct	tcc	agc	ctc	533
30	Tyr	Phe	Lys	Arg	Gln	Gly	Phe	Ser	Tyr	Gln	Gly	Arg	Ala	Ser	Ser	Leu	
				140					145					150			
	gat	gac	tta	cga	gcc	ctc	aag	ggg	aag	cat	ttt	ctc	ctg	gac	ccc	ttg	581
	Asp	Asp	Leu	Arg	Ala	Leu	Lys	Gly	Lys	His	Phe	Leu	Leu	Asp	Pro	Leu	
				155				160					165				
35	atg	ccg	gag	ctg	ctg	gtg	ttt	ccc	gcc	cag	aca	gat	ctg	cat	gaa	cac	629
	Met	Pro	Glu	Leu	Leu	Val	Phe	Pro	Ala	Gln	Thr	Asp	Leu	His	Glu	His	
				170			175					180					
	cca	ctg	tac	cgg	gcc	gga	cac	ctc	att	ctg	cag	gac	agg	gcc	agc	tgt	677
	Pro	Leu	Tyr	Arg	Ala	Gly	His	Leu	Ile	Leu	Gln	Asp	Arg	Ala	Ser	Cys	
40	185					190					195					200	
	ctc	cca	gcc	atg	ctg	ctg	gac	ccc	ccg	cca	ggc	tcc	cat	gtc	atc	gat	725
	Leu	Pro	Ala	Met	Leu	Leu	Asp	Pro	Pro	Pro	Gly	Ser	His	Val	Ile	Asp	
					205					210					215		
	gcc	tgt	gcc	gcc	cca	ggc	aat	aag	acc	agt	cac	ttg	gct	gct	ctt	ctg	773
45	Ala	Cys	Ala	Ala	Pro	Gly	Asn	Lys	Thr	Ser	His	Leu	Ala	Ala	Leu	Leu	
					220				225					230			
	aag	aac	caa	ggg	aag	atc	ttt	gcc	ttt	gac	ctg	gat	gcc	aag	cgg	ctg	821
	Lys	Asn	Gln	Gly	Lys	Ile	Phe	Ala	Phe	Asp	Leu	Asp	Ala	Lys	Arg	Leu	
				235				240					245				
50	gca	tcc	atg	gcc	acg	ctg	ctg	gcc	cgg	gct	ggc	gtc	tct	tgc	tgt	gaa	869
	Ala	Ser	Met	Ala	Thr	Leu	Leu	Ala	Arg	Ala	Gly	Val	Ser	Cys	Cys	Glu	
				250			255					260					
	ctg	gct	gag	gag	gac	ttc	ctg	gcg	gtc	tcc	ccc	tcg	gat	cca	cgc	tac	917
	Leu	Ala	Glu	Glu	Asp	Phe	Leu	Ala	Val	Ser	Pro	Ser	Asp	Pro	Arg	Tyr	
55	265					270					275					280	
	cat	gag	gtc	cac	tac	atc	ctg	ctg	gat	cct	tcc	tgc	agt	ggc	tcg	ggg	965
	His	Glu	Val	His	Tyr	Ile	Leu	Leu	Asp	Pro	Ser	Cys	Ser	Gly	Ser	Gly	
					285					290					295		
	atg	ccg	agc	aga	cag	ctg	gag	gag	ccc	ggg	gca	ggc	aca	cct	agc	ccg	1013
60	Met	Pro	Ser	Arg	Gln	Leu	Glu	Glu	Pro	Gly	Ala	Gly	Thr	Pro	Ser	Pro	
				300					305					310			
	gtg	cgt	ctg	cat	gcc	ctg	gca	ggc	ttc	cag	cag	cga	gcc	ctg	tgc	cac	1061
	Val	Arg	Leu	His	Ala	Leu	Ala	Gly	Phe	Gln	Gln	Arg	Ala	Leu	Cys	His	

	315	320	325	
	gcg ctc act ttc cct tcc ctg cag cgg ctc gtc tac tcc acg tgc tcc			1109
	Ala Leu Thr Phe Pro Ser Leu Gln Arg Leu Val Tyr Ser Thr Cys Ser			
	330	335	340	
5	ctc tgc cag gag gag aat gaa gac gtg gtg cga gat gcg ctg cag cag			1157
	Leu Cys Gln Glu Glu Asn Glu Asp Val Val Arg Asp Ala Leu Gln Gln			
	345	350	355	360
	aac ccg ggc gcc ttc agg cta gct ccc gcc ctg cct gcc tgg ccc cac			1205
	Asn Pro Gly Ala Phe Arg Leu Ala Pro Ala Leu Pro Ala Trp Pro His			
10		365	370	375
	cga ggc ctg agc acg ttc ccg ggt gcc gag cac tgc ctc cgg gcc tcc			1253
	Arg Gly Leu Ser Thr Phe Pro Gly Ala Glu His Cys Leu Arg Ala Ser			
	380	385	390	
	cct gag acc aca ctc agc agt ggc ttc ttc gtt gct gta att gaa cgg			1301
15	Pro Glu Thr Thr Leu Ser Ser Gly Phe Phe Val Ala Val Ile Glu Arg			
	395	400	405	
	gtc gag gtg cca agc tca gcc tca cag gcc aaa gca tca gca cca gaa			1349
	Val Glu Val Pro Ser Ser Ala Ser Gln Ala Lys Ala Ser Ala Pro Glu			
	410	415	420	
20	cgc aca ccc agc cca gcc cca aag aga aag aag aga cag caa aga gcc			1397
	Arg Thr Pro Ser Pro Ala Pro Lys Arg Lys Lys Arg Gln Gln Arg Ala			
	425	430	435	440
	gca gcc ggt gct tgc aca ccg cct tgc aca tagcagaggc tccgggctga			1447
	Ala Ala Gly Ala Cys Thr Pro Pro Cys Thr			
25		445	450	
	ctccttcttg gtgggaaagg aagatgcctg tcctctccgt ggaggaccct gggccctcac			1507
	cgcaggaagc agtttggtt ttgaaagggtt attgggtccc ttccttgggc tgtgttcttg			1567
	ctggtgagca aagtgttacc tgcaaaaata aaatgcagaa cgtactctac gacaaaaaaa			1627
	aaaaaaaaa			1637
30	<210> 71 <211> 1636 <212> DNA <213> Homo sapiens			
35	<220> <221> CDS <222> 30..1175			
40	<220> <221> sig_peptide <222> 30..77 <223> Von Heijne matrix score 3.71064775937629			
45	seq YAAAAGVLAVES/RQ			
	<400> 71			
	ctaatacgaaa agtaaaggcg cgcgggaac atg ggg ctg tat gct gca gct gca			53
		Met Gly Leu Tyr Ala Ala Ala Ala		
50		-15	-10	
	ggc gtg ttg gcc ggc gtg gag agc cgc cag ggc tct atc aag ggg ttg			101
	Gly Val Leu Ala Gly Val Glu Ser Arg Gln Gly Ser Ile Lys Gly Leu			
	-5	1	5	
	gtg tac tcc agc aac ttc cag aac gtg aag cag ctg tac gcg ctg gtg			149
55	Val Tyr Ser Ser Asn Phe Gln Asn Val Lys Gln Leu Tyr Ala Leu Val			
	10	15	20	
	tgc gaa acg cag cgc tac tcc gcc gtg ctg gat gct gtg atc gcc agc			197
	Cys Glu Thr Gln Arg Tyr Ser Ala Val Leu Asp Ala Val Ile Ala Ser			
	25	30	35	40
60	gcc ggc ctc ctc cgt gcg gag aag aag ctg cgg ccg cac ctg gcc aag			245
	Ala Gly Leu Leu Arg Ala Glu Lys Lys Leu Arg Pro His Leu Ala Lys			
	45	50	55	
	gtg cta gtg tat gag ttg ttg ttg gga aag ggc ttt cga ggg ggt ggg			293

	Val	Leu	Val	Tyr	Glu	Leu	Leu	Leu	Gly	Lys	Gly	Phe	Arg	Gly	Gly	Gly	
				60					65					70			
	ggc	cga	tgg	aag	gct	ctg	ttg	ggc	cgg	cac	cag	gcg	agg	ctc	aag	gct	341
5	Gly	Arg	Trp	Lys	Ala	Leu	Leu	Gly	Arg	His	Gln	Ala	Arg	Leu	Lys	Ala	
			75					80					85				
	gag	ttg	gct	cgg	ctc	aag	gtt	cat	cgg	ggt	gtg	agc	cgg	aat	gag	gac	389
	Glu	Leu	Ala	Arg	Leu	Lys	Val	His	Arg	Gly	Val	Ser	Arg	Asn	Glu	Asp	
			90				95					100					
	ctg	ttg	gaa	gtg	gga	tcc	agg	cct	ggt	cca	gcc	tcc	cag	ctg	cct	cga	437
10	Leu	Leu	Glu	Val	Gly	Ser	Arg	Pro	Gly	Pro	Ala	Ser	Gln	Leu	Pro	Arg	
							110				115					120	
	ttt	gtg	cgt	gtg	aac	act	ctc	aag	acc	tgc	tcc	gat	gat	gta	gtt	gat	485
	Phe	Val	Arg	Val	Asn	Thr	Leu	Lys	Thr	Cys	Ser	Asp	Asp	Val	Val	Asp	
					125					130					135		
15	tat	ttc	aag	aga	caa	ggg	ttc	tcc	tat	cag	ggt	cgg	gct	tcc	agc	ctc	533
	Tyr	Phe	Lys	Arg	Gln	Gly	Phe	Ser	Tyr	Gln	Gly	Arg	Ala	Ser	Ser	Leu	
					140				145					150			
	gat	gac	tta	cga	gcc	ctc	aag	ggg	aag	cat	ttt	ctc	ctg	gac	ccc	ttg	581
20	Asp	Asp	Leu	Arg	Ala	Leu	Lys	Gly	Lys	His	Phe	Leu	Leu	Asp	Pro	Leu	
			155					160					165				
	atg	ccg	gag	ctg	ctg	gtg	ttt	ccc	gcc	cag	aca	gat	ctg	cat	gaa	cac	629
	Met	Pro	Glu	Leu	Leu	Val	Phe	Pro	Ala	Gln	Thr	Asp	Leu	His	Glu	His	
			170				175					180					
	cca	ctg	tac	cgg	gcc	gga	cac	ctc	att	ctg	cag	gac	agg	gcc	agc	tgt	677
25	Pro	Leu	Tyr	Arg	Ala	Gly	His	Leu	Ile	Leu	Gln	Asp	Arg	Ala	Ser	Cys	
						190					195					200	
	ctc	cca	gcc	atg	ctg	ctg	gac	ccc	ccg	cca	ggc	tcc	cat	gtc	atc	gat	725
	Leu	Pro	Ala	Met	Leu	Leu	Asp	Pro	Pro	Gly	Ser	His	Val	Ile	Asp		
					205				210					215			
30	gcc	tgt	gcc	gcc	cca	ggc	aat	aag	acc	agt	cac	ttg	gct	gct	ctt	ctg	773
	Ala	Cys	Ala	Ala	Pro	Gly	Asn	Lys	Thr	Ser	His	Leu	Ala	Ala	Leu	Leu	
					220				225					230			
	aag	aac	caa	ggg	aag	atc	ttt	gcc	ttt	gac	ctg	gat	gcc	aag	cgg	ctg	821
35	Lys	Asn	Gln	Gly	Lys	Ile	Phe	Ala	Phe	Asp	Leu	Asp	Ala	Lys	Arg	Leu	
			235					240					245				
	gca	tcc	atg	gcc	acg	ctg	ctg	gcc	cgg	gct	ggc	gtc	tct	tgc	tgt	gaa	869
	Ala	Ser	Met	Ala	Thr	Leu	Leu	Ala	Arg	Ala	Gly	Val	Ser	Cys	Cys	Glu	
			250				255					260					
	ctg	gct	gag	gag	gac	ttc	ctg	gcg	gtc	tcc	ccc	tgc	gat	cca	cgc	tac	917
40	Leu	Ala	Glu	Glu	Asp	Phe	Leu	Ala	Val	Ser	Pro	Ser	Asp	Pro	Arg	Tyr	
							270				275					280	
	cat	gag	gtc	cac	tac	atc	ctg	ctg	gat	cct	tcc	tgc	agt	ggc	tgc	ggg	965
	His	Glu	Val	His	Tyr	Ile	Leu	Leu	Asp	Pro	Ser	Cys	Ser	Gly	Ser	Gly	
					285				290					295			
45	atg	ccg	agc	aga	cag	ctg	gag	gag	ccc	ggg	gca	ggc	aca	cct	agc	ccg	1013
	Met	Pro	Ser	Arg	Gln	Leu	Glu	Glu	Pro	Gly	Ala	Gly	Thr	Pro	Ser	Pro	
				300					305					310			
	gtg	cgt	ctg	cat	gcc	ctg	gca	gct	tcc	agc	agc	gag	ccc	tgt	gcc	acg	1061
50	Val	Arg	Leu	His	Ala	Leu	Ala	Ala	Ser	Ser	Ser	Glu	Pro	Cys	Ala	Thr	
					315				320				325				
	cgc	tca	ctt	tcc	ctt	ccc	tgc	agc	ggc	tgc	tct	act	cca	cgt	gct	ccc	1109
	Arg	Ser	Leu	Ser	Leu	Pro	Cys	Ser	Gly	Ser	Ser	Thr	Pro	Arg	Ala	Pro	
					330				335				340				
	tct	gcc	agg	agg	aga	atg	aag	acg	tgg	tgc	gag	atg	cgc	tgc	agc	aga	1157
55	Ser	Ala	Arg	Arg	Arg	Met	Lys	Thr	Trp	Cys	Glu	Met	Arg	Cys	Ser	Arg	
						345					350					360	
	acc	cgg	gcg	cct	tca	ggc	tagctcccgc	cctgacctgcc	tgccccccacc								1205
	Thr	Arg	Ala	Pro	Ser	Gly											
						365											
60	gaggcctgag	cacgttccccg	ggtgccgagc	actgcctccg	ggcctcccct	gagaccacac											1265
	tcagcagtgg	cttcttcggt	gctgtaattg	aacgggtcga	ggtgcccaagc	tcagcctcac											1325
	aggccaaagc	atcagcacca	gaacgcacac	ccagcccagc	cccaaagaga	aagaagagac											1385
	agcaaagagc	cgcagccggt	gcttgacac	cgcttgcac	atagcagagg	ctccgggctg											1445

```

actccttcct ggtgggaaag gaagatgcct gtcctctccg tggaggaccc tgggccctca 1505
ccgcaggaag cagtttgggt tttgaaaggt tattgggtcc cttccttggg ctgtgttctt 1565
gctgggtgagc aaagtgttac ctgcaaaaat aaaatgcaga acgtactcta cgacaaaaaa 1625
aaaaaaaaa a 1636
5
<210> 72
<211> 1758
<212> DNA
<213> Homo sapiens
10
<220>
<221> CDS
<222> 66..839
15
<220>
<221> sig_peptide
<222> 66..173
<223> Von Heijne matrix
      score 4.89555877630516
20      seq LLLRLNDAAALRA/LQ

<400> 72
agaggagggtg gcggtggtgg ccctcgccctg tggcccccgt gctgcttgca ctcgaactcg      60
tcgcc atg gag gag ctc cag gag cct ctg aga gga gag ctc cgg ctc tgc      110
25      Met Glu Glu Leu Gln Glu Pro Leu Arg Gly Glu Leu Arg Leu Cys
          -35          -30          -25
      ttc acg caa gct gcc cgg act agc ctc tta ctg ctc agg ctc aac gac      158
      Phe Thr Gln Ala Ala Arg Thr Ser Leu Leu Leu Arg Leu Asn Asp
          -20          -15          -10
30      gct gcc ctg cgg gcg ctg caa gag tgt cag cgg caa cag gta cgg ccg      206
      Ala Ala Leu Arg Ala Leu Gln Glu Cys Gln Arg Gln Gln Val Arg Pro
          -5          1          5          10
      gtg att gct ttc caa ggc cac cga ggg tat ctg aga ctc cca ggc cct      254
      Val Ile Ala Phe Gln Gly His Arg Gly Tyr Leu Arg Leu Pro Gly Pro
          15          20          25
35      ggt tgg tcc tgc ctc ttc tcc ttc ata gtg tcc cag tgt tgt cag gag      302
      Gly Trp Ser Cys Leu Phe Ser Phe Ile Val Ser Gln Cys Cys Gln Glu
          30          35          40
      ggc gct ggt ggt agc ttg gac ctt gtg tgc caa cgc ttc ctc agg tct      350
40      Gly Ala Gly Gly Ser Leu Asp Leu Val Cys Gln Arg Phe Leu Arg Ser
          45          50          55
      ggg cct aac agc ctc cac tgc ctg ggc tca ctc agg gag cgc ctc att      398
      Gly Pro Asn Ser Leu His Cys Leu Gly Ser Leu Arg Glu Arg Leu Ile
          60          65          70          75
45      att tgg gca gcc atg gat tct atc cca gcc cca tca tca gtt cag gga      446
      Ile Trp Ala Ala Met Asp Ser Ile Pro Ala Pro Ser Ser Val Gln Gly
          80          85          90
      cac aac ctg act gaa gat gcc aga cat cct gag agt tgg cag aac aca      494
      His Asn Leu Thr Glu Asp Ala Arg His Pro Glu Ser Trp Gln Asn Thr
          95          100          105
50      gga ggc tat tct gaa gga gat gca gta tca cag cca cag atg gca cta      542
      Gly Gly Tyr Ser Glu Gly Asp Ala Val Ser Gln Pro Gln Met Ala Leu
          110          115          120
      gag gag gtg tca gtg tca gat cca ctg gca agc aac caa gga cag tca      590
55      Glu Glu Val Ser Val Ser Asp Pro Leu Ala Ser Asn Gln Gly Gln Ser
          125          130          135
      ctc cca gga tcc tca agg gag cac atg gca cag tgg gaa gtg aga agc      638
      Leu Pro Gly Ser Ser Arg Glu His Met Ala Gln Trp Glu Val Arg Ser
          140          145          150          155
60      cag acc cat gtt cca aac aga gaa cct gtt cag gca ctg cct tcc tct      686
      Gln Thr His Val Pro Asn Arg Glu Pro Val Gln Ala Leu Pro Ser Ser
          160          165          170
      gcc agc cgg aaa cgt ctg gac aag aaa cgt tca gtg cct gta gcc act      734

```

Ala Ser Arg Lys Arg Leu Asp Lys Lys Arg Ser Val Pro Val Ala Thr
 175 180 185
 gta gaa ctg gaa gaa aag agg ttc aga act ctg cct tta gtg ccc ccc 782
 Val Glu Leu Glu Glu Lys Arg Phe Arg Thr Leu Pro Leu Val Pro Pro
 5 190 195 200
 cct aca agg cct gac caa tca gga ttt aca aga ggg aga aga ttg gga 830
 Pro Thr Arg Pro Asp Gln Ser Gly Phe Thr Arg Gly Arg Arg Leu Gly
 205 210 215
 gca aga aga tgaggacatg gaccccagat tagaacacaa ttcctcagtt 879
 10 Ala Arg Arg
 220
 caagaagatt ctgaatcccc aagtcctgaa gatataccag actacctcct gcaatacagg 939
 gccatccaca gtgcagaaca gcaacatgcc tatgagcagg actttgagac agattatgct 999
 gaataccgca tcctgcatgc ccgtgttggg actgcaagcc aaagggttcat agagctggga 1059
 15 gcagagatta aaagagtctg gcgaggaact ccagaatata aggtcctgga agacaagata 1119
 atccaggaat ataaaaagtt caggaagcag taccacagtt acagagaaga aaagcgtcgc 1179
 tgtgagtacc ttcaccagaa attgtcccac attaaagggtc tcatcctgga gtttgaggaa 1239
 aagaacaggg gcagctgaag ttatcaaggg aattttttgag cctctgctta gtgaaacaca 1299
 aaggaacaaa gcagctataa actaaataga atgcaactat ctgcttttct tatgctgacc 1359
 20 actggagtcc atgggtggcaa gtagagagct gctctagggtt cttgagggtt gggttttcatt 1419
 attaatTTTT aggggtatggg cactgtgcaa agactccata gctgtgccta ggagtctagg 1479
 aaaagtgaca gaggcttggc ttttttacct ttagttcagc caagtcattt tcaagtcctg 1539
 agaaatgaca tcattcttcag gataaaataa tgaggacatt agacaaacca aactaagtga 1599
 atttttagcct ggtagcctct ctaaggaaac agtaataata acttctgata agagttaaaa 1659
 25 gaacttgtag catacctgga tataacggga aagggcctgg gtgttaccga tgtactgaaa 1719
 atgaactttt accaacatgg ccaaaaaaaaaa aaaaaaaaaa 1758

<210> 73
 <211> 1647
 30 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 35 <222> 64..903

<220>
 <221> sig_peptide
 <222> 64..162
 40 <223> Von Heijne matrix
 score 10.6748773272319
 seq LLLLPFLPLLLLA/AP

<400> 73
 45 agctcaaggg gcctcgagga ctctctgcgt ctctggagac aagggcacta cacgcacttc 60
 aga atg aag agt tgc ggg agc atg ctg ggg ctc tgg ggg cag cgg ctc 108
 Met Lys Ser Cys Gly Ser Met Leu Gly Leu Trp Gly Gln Arg Leu
 -30 -25 -20
 ccc gcg gcg tgg gtc ctg ctt ctg ttg cct ttc ctg ccg ctg ctg ctg 156
 50 Pro Ala Ala Trp Val Leu Leu Leu Leu Pro Phe Leu Pro Leu Leu Leu
 -15 -10 -5
 ctt gca gcc ccc gcg ccc cac cgc gcg tcc tac aag ccg gtc atc gtg 204
 Leu Ala Ala Pro Ala Pro His Arg Ala Ser Tyr Lys Pro Val Ile Val
 1 5 10
 55 gtg cat ggg ctc ttc gac agc tcg tac agc ttc cgc cac ctg ctg gaa 252
 Val His Gly Leu Phe Asp Ser Ser Tyr Ser Phe Arg His Leu Leu Glu
 15 20 25 30
 tac atc aat gag aca cac ccc ggg act gtg gtg aca gtg ctc gat ctc 300
 Tyr Ile Asn Glu Thr His Pro Gly Thr Val Val Thr Val Leu Asp Leu
 35 40 45
 60 ttc gat ggg aga gag agc ttg cga ccc ctg tgg gaa cag gtg caa ggg 348
 Phe Asp Gly Arg Glu Ser Leu Arg Pro Leu Trp Glu Gln Val Gln Gly
 50 55 60

	ttc cga gag gct gtg gtc ccc atc atg gca aag gcc cct caa ggg gtg	396
	Phe Arg Glu Ala Val Val Pro Ile Met Ala Lys Ala Pro Gln Gly Val	
	65 70 75	
5	cat ctc atc tgc tac tcg cag ggg ggc ctt gtg tgc cgg gct ctg ctt	444
	His Leu Ile Cys Tyr Ser Gln Gly Gly Leu Val Cys Arg Ala Leu Leu	
	80 85 90	
	tct gtc atg gat gat cac aac gtg gat tct ttc atc tcc ctc tcc tct	492
	Ser Val Met Asp Asp His Asn Val Asp Ser Phe Ile Ser Leu Ser Ser	
	95 100 105 110	
10	cca cag atg gga cag tat gga gac acg gac tac ttg aag tgg ctg ttc	540
	Pro Gln Met Gly Gln Tyr Gly Asp Thr Asp Tyr Leu Lys Trp Leu Phe	
	115 120 125	
	ccc acc tcc atg cgg tct aac ctc tat cgg atc tgc tat agc ccc ctg	588
	Pro Thr Ser Met Arg Ser Asn Leu Tyr Arg Ile Cys Tyr Ser Pro Leu	
15	130 135 140	
	atc aat ggg gaa aga gac cat ccc aat gcc aca gta tgg cgg aag aac	636
	Ile Asn Gly Glu Arg Asp His Pro Asn Ala Thr Val Trp Arg Lys Asn	
	145 150 155	
20	ttt ctg cgt gtg ggc cac ctg gtg ctg att ggg ggc cct gat gat ggt	684
	Phe Leu Arg Val Gly His Leu Val Leu Ile Gly Gly Pro Asp Asp Gly	
	160 165 170	
	gtt att act ccc tgg cag tcc agc ttc ttt ggt ttc tat gat gca aat	732
	Val Ile Thr Pro Trp Gln Ser Ser Phe Phe Gly Phe Tyr Asp Ala Asn	
	175 180 185 190	
25	gag acc gtc ctg gag atg gag gag caa ctg gtt tat ctg cgg gat tct	780
	Glu Thr Val Leu Glu Met Glu Glu Gln Leu Val Tyr Leu Arg Asp Ser	
	195 200 205	
	ttt ggg ttg aag act cta ttg gcc cgg ggg gcc ata gtg agg tgt cca	828
	Phe Gly Leu Lys Thr Leu Leu Ala Arg Gly Ala Ile Val Arg Cys Pro	
30	210 215 220	
	atg gcc ggt atc tcc cac aca gcc tgg cac tcc aac cgt acc ctt tat	876
	Met Ala Gly Ile Ser His Thr Ala Trp His Ser Asn Arg Thr Leu Tyr	
	225 230 235	
35	gag acc tgc att gaa cct tgg ctc tcc tgaggatata ttcaggggtc	923
	Glu Thr Cys Ile Glu Pro Trp Leu Ser	
	240 245	
	cccaggaact cctcgggtcca gagaccaagt ggtggccttg gaaagcagat gtcaggcttt	983
	ggtgtgcctg tgaccacctc attgctccca tattatcccc catttttagt agagacgggg	1043
	ttttagtaga gacttggcct ccagaaaccc ccttcctctg ctccctccatg aatgacaatt	1103
40	ccaggcctcc cctacctcat gtcctctcat ttgggggatt gctccgtgct gtccctttct	1163
	ctcaaggccg aagttcggaa gtgagaaacc atgttttttaa cttgtggctg ctcttgctgc	1223
	tgctgtcctt ccgtatctgg ctgtatgggt ggagaaccca cccactgccc accacagggg	1283
	tctccttcca ggccactcag gacattttta gcttctctcc tccccatgtt cccttttttc	1343
	tctaaagtcc cctgacatca gccctcccaa ctccctaagag ggactaccca tgagagtggg	1403
45	gttctgaggg tcccctatgg ggacagttcc gttcttgaag tgtcagtgtt gggaatatc	1463
	tgtggcctat gagggccatc tcaggtttgg ggatccccca gtccctatga tcagtgttgg	1523
	agtaccccc tgaggagagcc tagtttctt gagggccccag gccctctttt aactaccttt	1583
	gaataggtgt tatccctgta tttatggaaa taaagttcca tttcctcaaa aaaaaaaaaa	1643
	aaaa	1647
50	<210> 74	
	<211> 1646	
	<212> DNA	
	<213> Homo sapiens	
55	<220>	
	<221> CDS	
	<222> 64..585	
60	<220>	
	<221> sig_peptide	
	<222> 64..162	
	<223> Von Heijne matrix	

score 10.6748773272319
seq LLLLPFLPLLLLA/AP

<400> 74
5 agctcaaggg gcctcgagga ctctctgcgt ctctggagac aagggcacta cacgcacttc 60
aga atg aag agt tgc ggg agc atg ctg ggg ctg tgg ggg cag cgg ctg 108
Met Lys Ser Cys Gly Ser Met Leu Gly Leu Trp Gly Gln Arg Leu
-30 -25 -20
ccc gcg gcg tgg gtc ctg ctt ctg ttg cct ttc ctg ccg ctg ctg ctg 156
10 Pro Ala Ala Trp Val Leu Leu Leu Leu Pro Phe Leu Pro Leu Leu Leu
-15 -10 -5
ctt gca gcc ccc gcg ccc cac cgc gcg tcc tac aag ccg gtc atc gtg 204
Leu Ala Ala Pro Ala Pro His Arg Ala Ser Tyr Lys Pro Val Ile Val
1 5 10
15 gtg cat ggg ctg ttc gac agc tcg tac agc ttc cgc cac ctg ctg gaa 252
Val His Gly Leu Phe Asp Ser Ser Tyr Ser Phe Arg His Leu Leu Glu
15 20 25 30
tac atc aat gag aca cac ccc ggg act gtg gtg aca gtg ctg gat ctg 300
Tyr Ile Asn Glu Thr His Pro Gly Thr Val Val Thr Val Leu Asp Leu
20 35 40 45
ttc gat ggg aga gag agc ttg cga ccc ctg tgg gaa cag gtg caa ggg 348
Phe Asp Gly Arg Glu Ser Leu Arg Pro Leu Trp Glu Gln Val Gln Gly
50 55 60
25 ttc cga gag gct gtg gtc ccc atc atg gca aag gcc cct caa ggg gtg 396
Phe Arg Glu Ala Val Val Pro Ile Met Ala Lys Ala Pro Gln Gly Val
65 70 75
cat ctg atc tgc tac tcg cag ggg ggc ctt gtg tgc cgg gct ctg ctt 444
His Leu Ile Cys Tyr Ser Gln Gly Gly Leu Val Cys Arg Ala Leu Leu
80 85 90
30 tct gtc atg gat gat cac aac gtg gat tct ttc atc tcc ctg tcc tct 492
Ser Val Met Asp Asp His Asn Val Asp Ser Phe Ile Ser Leu Ser Ser
95 100 105 110
cca cag atg gga cag tat gga gac acg gac tac ttg aag tgg ctg ttc 540
Pro Gln Met Gly Gln Tyr Gly Asp Thr Asp Tyr Leu Lys Trp Leu Phe
115 120 125
35 ccc acc tcc atg cgg tct aac ctg tat cgg atc tgc tat agc ccc 585
Pro Thr Ser Met Arg Ser Asn Leu Tyr Arg Ile Cys Tyr Ser Pro
130 135 140
tgatcaatgg ggaaagagac catcccaatg ccacagtatg gcggaagaac tttctgcgtg 645
40 tggggccacct ggtgctgatt gggggccctg atgatggtgt tattactccc tggcagttcca 705
gcttcttttgg tttctatgat gcaaatagaga ccgtcctgga gatggaggag caactgggtt 765
atctgcggga tttcttttggg ttgaagactc tattggcccc gggggccata gtgaggtgtc 825
caatggccgg tatctcccac acagcctggc actccaaccg taccctttat gagacctgca 885
ttgaaccttg gctctcctga ggatatattc aggggtcccc aggaactcct cgggtccagag 945
45 accaagtggg ggccttgga agcagatgtc aggcctttggg gtgcctgtga ccacctcatt 1005
gtcctccat tatccccat ttttagtaga gacgggggtt tagtagagac ttggcctccc 1065
agaaccccc tctctgtctc ctccatgaat gacaattcca ggctccccct acctcatgtc 1125
ctctcatttg ggggattgct ccgtgctgtc cttttctctc aaggccgaag ttcggaagtg 1185
agaaaccatg tttttaactt gtggctgtc ttgctgctgc tgctcctccg tatctggctg 1245
50 tatgggtgga gaaccacacc actgcccacc acaggggtct cctccaggc cactcaggac 1305
atcttttagct tctctcctcc ccatgttccc tttttctct aaagtcccc gacatcagcc 1365
ctcccaactc ctaagaggga ctacctatga gagtggggtt ctgaggctcc cctatgggga 1425
cagttccgtt cttgaagtgt cagtgttggg gaatatctgt ggccatagag gcccatctca 1485
ggtttgggga tccccagtc cctatgatca gtgttgaggt accccctgg gagagcctag 1545
55 tttctttgag gccccaggc ctcttttaac tacctttgaa taggtgttat cctgtattt 1605
atggaaataa agttccattt cctcaaaaaa aaaaaaaaaa a 1646

<210> 75

<211> 1963

60 <212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 274..753

<220>

5 <221> sig_peptide

<222> 274..324

<223> Von Heijne matrix

score 4.4969823290892

seq FAAFCYMLSLVLC/AA

10

<400> 75

```

cttcttcgat ttgcggacgg ttccctccag cgactctcga cacacgtttt cctgtcttcg      60
ccggagggcc gggctctggg tgcgcggagc ctgcgggaat ccagcgctta ttcgctaacc      120
ctcgagtcgc ttcgctagct gtgcgccttc ctgggcacta gcctggagag gagcgtgcag      180
15 acgcggctcc ttggagggag tgcggtcctc tagggaggca tcgggctcct aggggcttct      240
tggcgtgtgt ggtgggattg gggtcgcgcg gcc atg gcc ttc act ttc gct gcg      294
                                     Met Ala Phe Thr Phe Ala Ala
                                     -15
ttc tgc tac atg ctg tct ctg gtg ctg tgc gct gcg ctc atc ttc ttc      342
20 Phe Cys Tyr Met Leu Ser Leu Val Leu Cys Ala Ala Leu Ile Phe Phe
   -10          -5          1          5
gcc atc tgg cac ata att gcc ttt gat gag tta agg aca gat ttt aag      390
Ala Ile Trp His Ile Ile Ala Phe Asp Glu Leu Arg Thr Asp Phe Lys
          10          15          20
25 agc ccc ata gac cag tgc aat cct gtt cat gcg agg gaa cgg ttg agg      438
Ser Pro Ile Asp Gln Cys Asn Pro Val His Ala Arg Glu Arg Leu Arg
          25          30          35
aac atc gag cgc atc tgc ttc ctt ctg cga aag ctg gtg ctg cca gaa      486
Asn Ile Glu Arg Ile Cys Phe Leu Leu Arg Lys Leu Val Leu Pro Glu
30          40          45          50
tac tcc atc cat agc ctc ttc tgc att atg ttc ctg tgt gcg caa gag      534
Tyr Ser Ile His Ser Leu Phe Cys Ile Met Phe Leu Cys Ala Gln Glu
55          60          65          70
tgg ctc acg ctg ggg ctg aat gtc cct cta ctt ttc tat cac ttc tgg      582
Trp Leu Thr Leu Gly Leu Asn Val Pro Leu Leu Phe Tyr His Phe Trp
          75          80          85
agg tat ttc cac tgt cca gca gat agc tca gaa cta gcc tac gac cca      630
Arg Tyr Phe His Cys Pro Ala Asp Ser Ser Glu Leu Ala Tyr Asp Pro
          90          95          100
40 ccg gtg gtc atg aat ccc gac act ttg agt tac tgt cag aag gag gcc      678
Pro Val Val Met Asn Pro Asp Thr Leu Ser Tyr Cys Gln Lys Glu Ala
          105          110          115
tgg tgt aag ctg gcc ttc tat ctc ctc tcc ttc ttc tac tac ctt tac      726
Trp Cys Lys Leu Ala Phe Tyr Leu Leu Ser Phe Phe Tyr Tyr Leu Tyr
45          120          125          130
tgc atg atc tac act tta gtg agc tct taacgcaaag accatgcaca      773
Cys Met Ile Tyr Thr Leu Val Ser Ser
          135          140
tcatcagaga ctgagatggg agaggcctga gacggagagg tgcatttctg ctggtgactg      833
50 gaggagggac cagaatgagg atacgtgaga aatagacccg gcaggcagtc agactgaatg      893
ggagctggaa tcacgcagca gttgggagcc gagttaaccc tgcgtgtctg tgtcaccctg      953
tttgtcaatc tttggcattc gaattccaca cacgggggtcc tagagccctt ctgagcatca      1013
gtggtgtggg ggagtaggtg acgaaacact agacctctcc tgagagagaa ttgctgcttc      1073
ctgaatccac ttcattgaac agcaccttgc aagttcaaat gagttcctgg gagcggaggc      1133
55 tggaaggcca caaggtgctt gctaaggaaac agaatagcc ccaagtcctgc      1193
agggacctgt tgaaagcctc gagaatgtct tggctgcccc agactcttgt tgcctttctt      1253
ccaagccatg gccatgccct ttttctcaaa tgggaggggc tggagggtgt gtgggatttg      1313
tcttcagctg caaccagcct tgagcctgct gggctatttt cagctgagga ggggtaatat      1373
aggaataatg catttttgaa acgtttgcaa catgatcaag gtgttagttc tccaccacac      1433
60 aagttgtatt cttcttttgc cacctcaaac catcacagag tctttaaatg caaatcaatt      1493
ggtcaatgct agtcaaagct atgttcttac aaaaacccca gacagctcag agctcagaaa      1553
atcctgtgga gtggctgctc tgtaccgtgg gcatccggca gccaggaagt gagacaacat      1613
aattataact ttgttttatg atgctgcac ctttgactg tttaggtcga cgtgaggaca      1673

```

```

tcattcttatt tagaattttc cgtttggcat tctcttttgg gtgggagtta tgctgggggt 1733
tgtaaataat gacaaggctg agatttttat gatgtttaaa ttgggcacaa tgattttgac 1793
cttattcccc aaacttcttt tcttttctac tgtttaacat acacaggcta tttatacacg 1853
tccccagctc ccatctgaaa cctgtgactc aggtttatga atggtgtttg tgtagcaaca 1913
5 cattgtgtgc tatgtttatt aaaatgcagc gacaaaaaaa aaaaaaaaaa 1963

<210> 76
<211> 1757
<212> DNA
10 <213> Homo sapiens

<220>
<221> CDS
<222> 191..1468
15 <220>
<221> sig_peptide
<222> 191..274
<223> Von Heijne matrix
20 score 4.02941490119842
seq GXLLEPFVHQVGG/HS

<400> 76
catttttggtg cgagagaaac aataggacgg aaacgccgag gaacccggct gaggcggcag 60
25 cagagcatcc tggccagaac aagccaagga gccaaagacga gagggacaca ctgacaaaca 120
acagacagaa gacgtactgg ccgctggact ccgctgcctc ccccatctcc ccgccatctg 180
cgccccggagg atg agc cca gcc ttc agg gcc atg gat gtg gag ccc cgc 229
Met Ser Pro Ala Phe Arg Ala Met Asp Val Glu Pro Arg
-25 -20
30 gcc aaa ggc gtc ctt ctg gag ccc ttt gtc cac cag gtc ggg ggg cac 277
Ala Lys Gly Val Leu Leu Glu Pro Phe Val His Gln Val Gly Gly His
-15 -10 -5 1
tca tgc gtg ctc cgc ttc aat gag aca acc ctg tgc aag ccc ctg gtc 325
Ser Cys Val Leu Arg Phe Asn Glu Thr Thr Leu Cys Lys Pro Leu Val
35 5 10 15
cca agg gaa cat cag ttc tac gag acc ctc cct tct gag atg cgc aaa 373
Pro Arg Glu His Gln Phe Tyr Glu Thr Leu Pro Ser Glu Met Arg Lys
20 25 30
40 ttc act ccc cag tac aaa ggt gtg gta tct gtg cgc ttt gaa gaa gat 421
Phe Thr Pro Gln Tyr Lys Gly Val Val Ser Val Arg Phe Glu Glu Asp
35 40 45
gaa gac agg aac ttg tgt cta ata gca tat cca ttg aaa ggg gac cat 469
Glu Asp Arg Asn Leu Cys Leu Ile Ala Tyr Pro Leu Lys Gly Asp His
50 55 60 65
45 gga att gtg gac att gta gat aat tca gac tgt gaa cca aaa agt aag 517
Gly Ile Val Asp Ile Val Asp Asn Ser Cys Glu Pro Lys Ser Lys
70 75 80
ctc cta agg tgg aca aca aac aaa aaa cat cat gtc tta gaa aca gaa 565
Leu Leu Arg Trp Thr Thr Asn Lys Lys His His Val Leu Glu Thr Glu
85 90 95
50 aag acc cct aag gac tgg gtg cgt cag cac cgt aaa gag gag aaa atg 613
Lys Thr Pro Lys Asp Trp Val Arg Gln His Arg Lys Glu Glu Lys Met
100 105 110
aag agc cat aag tta gaa gaa gaa ttt gag tgg cta aag aaa tct gaa 661
Lys Ser His Lys Leu Glu Glu Phe Glu Trp Leu Lys Lys Ser Glu
115 120 125
gtc ttg tac tac act gta gag aag aag ggg aat ata agt tcc cag ctt 709
Val Leu Tyr Tyr Thr Val Glu Lys Lys Gly Asn Ile Ser Ser Gln Leu
130 135 140 145
60 aaa cac tat aac cct tgg agc atg aaa tgt cac cag caa cag tta cag 757
Lys His Tyr Asn Pro Trp Ser Met Lys Cys His Gln Gln Gln Leu Gln
150 155 160
aga atg aag gag aat gca aag cat cgg aac cag tac aaa ttt atc tta 805

```

```

    Arg Met Lys Glu Asn Ala Lys His Arg Asn Gln Tyr Lys Phe Ile Leu
      165      170      175
    ctg gaa aac ctg act tcc cgc tat gag gtg cct tgt gtc ctt gac ctc      853
    Leu Glu Asn Leu Thr Ser Arg Tyr Glu Val Pro Cys Val Leu Asp Leu
      180      185      190
5   aag atg ggc aca cga caa cat ggt gat gat gct tca gag gag aag gca      901
    Lys Met Gly Thr Arg Gln His Gly Asp Asp Ala Ser Glu Glu Lys Ala
      195      200      205
10  gcc aac cag atc cga aaa tgt cag cag agc aca tct gca gtc att ggt      949
    Ala Asn Gln Ile Arg Lys Cys Gln Gln Ser Thr Ser Ala Val Ile Gly
      210      215      220      225
    gtg cgt gtg tgt ggc atg cag gtg tac caa gca ggc agt ggg cag ctc      997
    Val Arg Val Cys Gly Met Gln Val Tyr Gln Ala Gly Ser Gly Gln Leu
      230      235      240
15  atg ttc atg aac aag tac cat gga cgg aag cta tcg gtg cag ggc ttc      1045
    Met Phe Met Asn Lys Tyr His Gly Arg Lys Leu Ser Val Gln Gly Phe
      245      250      255
    aag gag gca ctt ttc cag ttc ttc cac aat ggg cgg tac ctg cgc cgt      1093
    Lys Glu Ala Leu Phe Gln Phe Phe His Asn Gly Arg Tyr Leu Arg Arg
      260      265      270
20  gaa ctc ctg ggc cct gtg ctc aag aag ctg act gag ctc aag gca gtg      1141
    Glu Leu Leu Gly Pro Val Leu Lys Lys Leu Thr Glu Leu Lys Ala Val
      275      280      285
    ttg gag cga cag gag tcc tac cgc ttc tac tca agc tcc ctg ctg gtc      1189
    Leu Glu Arg Gln Glu Ser Tyr Arg Phe Tyr Ser Ser Ser Leu Leu Val
      290      295      300      305
    att tat gat ggc aag gag cgg ccc gaa gtg gtc ctg gac tca gat gct      1237
    Ile Tyr Asp Gly Lys Glu Arg Pro Glu Val Val Leu Asp Ser Asp Ala
      310      315      320
30  gag gat ttg gag gac ctg tca gag gaa tca gct gat gag tct gct ggt      1285
    Glu Asp Leu Glu Asp Leu Ser Glu Glu Ser Ala Asp Glu Ser Ala Gly
      325      330      335
    gcc tat gcc tac aaa ccc atc ggc gcc agc tct gta gat gtg cgc atg      1333
    Ala Tyr Ala Tyr Lys Pro Ile Gly Ala Ser Ser Val Asp Val Arg Met
      340      345      350
35  atc gac ttt gca cac acc acc tgc agg ctg tat ggc gag gac acc gtg      1381
    Ile Asp Phe Ala His Thr Thr Cys Arg Leu Tyr Gly Glu Asp Thr Val
      355      360      365
40  gtg cat gag ggc cag gat gct ggc tat atc ttc ggg ctc cag agc ctg      1429
    Val His Glu Gly Gln Asp Ala Gly Tyr Ile Phe Gly Leu Gln Ser Leu
      370      375      380      385
    ata gac att gtc aca gag ata agt gag gag agt ggg gag tgagcttgct      1478
    Ile Asp Ile Val Thr Glu Ile Ser Glu Glu Ser Gly Glu
      390      395
45  agctgctcca gtacttgaga ggcactctgt gtcccaggca cagctgtgct gcgtcaggga      1538
    ggaagccagt atggccaggt ggtggctcct gcagcctgga gctgatgtgc agtggcctct      1598
    gtgagcccca gctgagcca gtcccagctg tgcttgaggt ctttatttat tttaactatt      1658
    tcttcaacat tccacatttg atgatgatac ctctttcttc cctgagtgta tatgttctaa      1718
    tacaaatctt tttgtttatt gtaaaaaaaaa aaaaaaaaaa      1757
50  <210> 77
    <211> 2027
    <212> DNA
    <213> Homo sapiens
55  <220>
    <221> CDS
    <222> 48..950
60  <220>
    <221> sig_peptide
    <222> 48..107
    <223> Von Heijne matrix

```

score 6.64507667657896
seq LLPLLSLLVGAWL/KL

<400> 77

```

5  atgcgcagcg gggccgtggg tgtacgcggc gcagcgcggc agtcctg atg gcc cgg      56
                                   Met Ala Arg
                                   -20
   cat ggg tta ccg ctg ctg ccc ctg ctg tgc ctc ctg gtc ggc gcg tgg      104
   His Gly Leu Pro Leu Leu Pro Leu Leu Ser Leu Leu Val Gly Ala Trp
10  -15                               -10                               -5
   ctc aag cta gga aat gga cag gct act agc atg gtc caa ctg cag ggt      152
   Leu Lys Leu Leu Gly Asn Gly Gln Ala Thr Ser Met Val Gln Leu Gln Gly
   1                               5                               10                               15
15  ggg aga ttc ctg atg gga aca aat tct cca gac agc aga gat ggt gaa      200
   Gly Arg Phe Leu Met Gly Thr Asn Ser Pro Asp Ser Arg Asp Gly Glu
                                   20                               25                               30
   ggg cct gtg cgg gag gcg aca gtg aaa ccc ttt gcc atc gac ata ttt      248
   Gly Pro Val Arg Glu Ala Thr Val Lys Pro Phe Ala Ile Asp Ile Phe
                                   35                               40                               45
20  cct gtc acc aac aaa gat ttc agg gat ttt gtc agg gag aaa aag tat      296
   Pro Val Thr Asn Lys Asp Phe Arg Asp Phe Val Arg Glu Lys Lys Tyr
                                   50                               55                               60
   cgg aca gaa gct gag atg ttt gga ttg agc ttt gtc ttt gag gac ttt      344
   Arg Thr Glu Ala Glu Met Phe Gly Leu Ser Phe Val Phe Glu Asp Phe
25  65                               70                               75
   gtc tct gat gag ctg aga aac aaa gcc acc cag cca atg aag tct gta      392
   Val Ser Asp Glu Leu Arg Asn Lys Ala Thr Gln Pro Met Lys Ser Val
   80                               85                               90                               95
30  ctc tgg tgg ctt cca gtg gaa aag gca ttt tgg agg cag cct gca ggt      440
   Leu Trp Trp Leu Pro Val Glu Lys Ala Phe Trp Arg Gln Pro Ala Gly
                                   100                               105                               110
   cct ggc tct ggc atc cga gag aga ctg gag cac cca gtg tta cac gtg      488
   Pro Gly Ser Gly Ile Arg Glu Arg Leu Glu His Pro Val Leu His Val
                                   115                               120                               125
35  agc tgg aat gac gcc cgt gcc tac tgt gct tgg cgg gga aaa cga ctg      536
   Ser Trp Asn Asp Ala Arg Ala Tyr Cys Ala Trp Arg Gly Lys Arg Leu
                                   130                               135                               140
   ccc acg gag gaa gag tgg gag ttt gcc gcc cga ggg ggc ttg aag ggt      584
   Pro Thr Glu Glu Glu Trp Glu Phe Ala Ala Arg Gly Gly Leu Lys Gly
40  145                               150                               155
   caa gtt tac cca tgg ggg aac tgg ttc cag cca aac cgc acc aac ctg      632
   Gln Val Tyr Pro Trp Gly Asn Trp Phe Gln Pro Asn Arg Thr Asn Leu
   160                               165                               170                               175
45  tgg cag gga aag ttc ccc aag gga gac aaa gct gag gat ggc ttc cat      680
   Trp Gln Gly Lys Phe Pro Lys Gly Asp Lys Ala Glu Asp Gly Phe His
                                   180                               185                               190
   gga gtc tcc cca gtg aat gct ttc ccc gcc cag aac aac tac ggg ctc      728
   Gly Val Ser Pro Val Asn Ala Phe Pro Ala Gln Asn Asn Tyr Gly Leu
                                   195                               200                               205
50  tat gac ctc ctg ggg aac gtg tgg gag tgg aca gca tca ccg tac cag      776
   Tyr Asp Leu Leu Gly Asn Val Trp Glu Trp Thr Ala Ser Pro Tyr Gln
                                   210                               215                               220
   gct gct gag cag gac atg cgc gtc ctc cgg ggg gca tcc tgg atc gac      824
   Ala Ala Glu Gln Asp Met Arg Val Leu Arg Gly Ala Ser Trp Ile Asp
55  225                               230                               235
   aca gct gat ggc tct gcc aat cac cgg gcc cgg gtc acc acc agg atg      872
   Thr Ala Asp Gly Ser Ala Asn His Arg Ala Arg Val Thr Thr Arg Met
   240                               245                               250                               255
60  ggc aac act cca gat tca gcc tca gac aac ctc ggt ttc cgc tgt gct      920
   Gly Asn Thr Pro Asp Ser Ala Ser Asp Asn Leu Gly Phe Arg Cys Ala
                                   260                               265                               270
   gca gac gca ggc cgg ccg cca ggg gag ctg taagcagccg ggtggtgaca      970
   Ala Asp Ala Gly Arg Pro Pro Gly Glu Leu

```

	275	280		
	aggagaaaag ccttctaggg tcaactgtcat tccctggcca tgttgcaaac agcgcaattc	1030		
	caagctcgag agcttcagcc tcaggaaaga acttcccctt ccctgtctcc catccctctg	1090		
	tggcaggcgc ctctcaccag ggcaggagag gactcagcct cctgtgtttt ggagaagggg	1150		
5	cccaatgtgt gttgacgatg gctggggggcc aggtgtttct gttagaggcc aagtattatt	1210		
	gacacaggat tgcaaacaca caaacaattg gaacagagca ctctgaaagg ccatttttta	1270		
	agcatttttaa aatctattct ctcccccttt ctccctggat gattcaggaa gctgacattg	1330		
	tttccctcaag gcagaatttt cctgggttctg ttttctcagc cagttgctgt ggaaggagaa	1390		
	tgctttcttt gtggcctcat ctgtggtttc gtgtccctct gaaggaaact agtttccact	1450		
10	gtgtaacagg cagacatgta actattttaa gcacagttca gtcctaaaag ggtctgggag	1510		
	aaccagatga tgtactaggt gaagcattgc attgtgggaa tcacaaagca aatagtactc	1570		
	cagaaagaca aatatcagaa gcttccctatt cttttttttt tttttttttt tttggagaca	1630		
	gggtctttct ctgttgccca ggctagagtg cactgggtgat cacggctcac tctagccttg	1690		
	aattcctggg cccaagcaat tctccacct cagcctcctg agtagctggg actacaagtg	1750		
15	tgcaccacca tgccctggcta attttttgaa tttttgtagt gatgggatct cgctctgttg	1810		
	cccagggtgg tctcgaactc ctggcctcaa gcgatcctcc cacctcgacc tcccaaagtg	1870		
	ctgggattac aggtgtgagc cacctcgctt gggccccctt ctccatatgc ctccaaaaac	1930		
	atgtccctgg agagtagcct gctccacacac tgtcactgga tgtcatgggg ccaataaaat	1990		
	ctcctgcaat tgtgtatctc aaaaaaaaaa aaaaaaa	2027		
20	<210> 78			
	<211> 1880			
	<212> DNA			
	<213> Homo sapiens			
25	<220>			
	<221> CDS			
	<222> 156..512			
30	<220>			
	<221> sig_peptide			
	<222> 156..206			
	<223> Von Heijne matrix			
	score 3.55618791452243			
35	seq WLTAVASLLPSPG/NS			
	<400> 78			
	atatacaggt ggcagctctc gtcccctgag agcggggcgaa ggccagggtc ccacactcgc	60		
	agcgtctcga cggcgtcgg gacctccctc gtccactgct tgagttccag aggtgggtgc	120		
40	ttccctgtcc tgaacttcag agtgcggagt cataa atg ggt tcc ggc tgg ctt	173		
	Met Gly Ser Gly Trp Leu			
	-15			
	act gca gta gcc tcg ctc ctc ccc agc ccc ggt aac tcc gag cta ccc	221		
	Thr Ala Val Ala Ser Leu Leu Pro Ser Pro Gly Asn Ser Glu Leu Pro			
45	-10 -5 1 5			
	gtc cag gcc ctc ggg cgt cgc ggg ggc agg gac tgg gcg cgg aac gag	269		
	Val Gln Ala Leu Gly Arg Arg Gly Gly Arg Asp Trp Ala Arg Asn Glu			
	10 15 20			
	gca ggg agg gac ctg gaa aaa cca ccc aga ttg cat tgc agt ggg cga	317		
50	Ala Gly Arg Asp Leu Glu Lys Pro Pro Arg Leu His Cys Ser Gly Arg			
	25 30 35			
	ggc cgc ctg gag gag ccg gtt ccc cct aac cac ctc ccc gtg ggg ctc	365		
	Gly Arg Leu Glu Glu Pro Val Pro Pro Asn His Leu Pro Val Gly Leu			
	40 45 50			
55	tcg gtg cgc ggt tcc cag gtg ctc agc tct gct ggg ccc agg agg tgc	413		
	Ser Val Arg Gly Ser Gln Val Leu Ser Ser Ala Gly Pro Arg Arg Cys			
	55 60 65			
	cgc ctc aca ggg acg cgg aac ccc gtg cgt ggc ccc cgc cgg gtg gaa	461		
	Arg Leu Thr Gly Thr Arg Asn Pro Val Arg Gly Pro Arg Arg Val Glu			
60	70 75 80 85			
	cag ata gcg cgg ggc ggt ccg gag gct cgt cgc caa gca ggt gac tct	509		
	Gln Ile Ala Arg Gly Gly Pro Glu Ala Arg Arg Gln Ala Gly Asp Ser			
	90 95 100			

	tgc tgaaaaagtg gttggaacac ttaaggaaac ccggccccgc ctgttctttc	562
	Cys	
	taggtctttg gagtttggat taatcatttg ttagcccggt ttggataaac cgaagacttt	622
	attaaatcag cgcgtttaac aggaattccg cagtagtata cacattagaa tcttgagtct	682
5	tggagttgaa catattcaca cagacttgcc ttcttctctgt ttagtttatg ccttggtgtc	742
	cggtattgga acgctaagct tgtgggagtt gtttacatcc tactgctcaa ggtcatcgct	802
	aagggtgat ttttcacaaa aagaatttgc aacctccggc atgaatgact taagggaagt	862
	ctaatacccg tttctgattt tttttttttt ttaatttaaa agttaatctt tctgggccgg	922
	gcgcggtggc tcacgcctgt aatcccagca ctttgggagg ccgagggcga tcacgaggtc	982
10	aggagttcga gaccagcctg accaacaatgg tgaaaccccg tctctactaa aaacacaaaa	1042
	attagccggg cggggtggcg cgcacctgta atcccagctg ctcggggagg tgaggcagga	1102
	gaatcgcttg aacctgggag gcgggggggt gcagtgagcc gagatctggc cattgcactc	1162
	cagcgtgggc aacagagtga gactccatct caaaaaaaaaa ggttaatctt tccaactaga	1222
	ttttcaagga tgaggatttt gttgttgttg ttgttgttgt tctcaaagt attcccagg	1282
15	cttggaacag agcctgacat atactaggca ctcaacaaat atttgttgaa tgattgtaat	1342
	gagtaacacc ctttttgca gatctttgtc ttctgagcct agggcatagg tcatcactgc	1402
	aggggtgaga ttgtcaaaat gggagcttac aggtaattta agacttaaat gtttaaagag	1462
	tatgtgctca ttcttcaaca aacttacttt tgtaaatta aaatggtaaa atgtggtgga	1522
	ggggttgga tatatgtaat tcaagacagt tctgaataca aaaatgtttt actgtctatc	1582
20	accaccatct ataaatctaa ttcactaagg ataactctgt taagggtggc ggaaagaacc	1642
	ttgaggagag aggccttattt aagtattggc tcaggaccac acctaaaatt ctcaaacgt	1702
	tgagattctg ttgttttgtt tttaagcgcc agagacccaa gttgaggaac agcctataaa	1762
	ataactggcc tgtactctta catacatgaa agccatcaaa gacaaagact gaagaagaac	1822
	ttttgcagat taaaggactt taagagacat gatcctgaac caaaaaaaaaa aaaaaaaaaa	1880
25	<210> 79	
	<211> 584	
	<212> DNA	
	<213> Homo sapiens	
30	<220>	
	<221> CDS	
	<222> 67..351	
35	<220>	
	<221> sig_peptide	
	<222> 67..183	
	<223> Von Heijne matrix	
	score 10.6473524146908	
40	seq FLCALCSFCPISA/AS	
	<400> 79	
	ctgattcttc gaaatgatata aagtcctgag ggcttcagtc ccattcgcgc actcatactt	60
	gcaatc atg gac tac agc cgt gtc ttt cag ggt gtg ttc ttc acc ttc	108
45	Met Asp Tyr Ser Arg Val Phe Gln Gly Val Phe Phe Thr Phe	
	-35 -30	
	aag cat gct ttt gct gat ggt gct tgg gat ctt tca ttt ctc tgt gct	156
	Lys His Ala Phe Ala Asp Gly Ala Trp Asp Leu Ser Phe Leu Cys Ala	
	-25 -20 -15 -10	
50	ctt tgc agt ttc tgc cca atc tca gct gcc tct ggc aga cct tac agg	204
	Leu Cys Ser Phe Cys Pro Ile Ser Ala Ala Ser Gly Arg Pro Tyr Arg	
	-5 1 5	
	tac ttg gaa ttc tgg aga tta tac ctg tct cct agt tcc atg gaa aat	252
	Tyr Leu Glu Phe Trp Arg Leu Tyr Leu Ser Pro Ser Ser Met Glu Asn	
55	10 15 20	
	gga gtt caa aaa ttc cac gaa act ttt ttc att gtc ttt ttg ctt ttg	300
	Gly Val Gln Lys Phe His Glu Thr Phe Phe Ile Val Phe Leu Leu Leu	
	25 30 35	
	ttt gat atc gag agg aaa gga aaa agt tct gtt tgt cca ttt tgt tac	348
60	Phe Asp Ile Glu Arg Lys Gly Lys Ser Ser Val Cys Pro Phe Cys Tyr	
	40 45 50 55	
	aga taaggaaagt ggtttcacaa aggttaagca acttggttcag tgttaccag	401
	Arg	

	caaagagcag	aatgattttc	aacatttcagt	ttaaaagtcg	gcggggggca	gtggctcaca	461
	cctgtaatat	cagcaacttg	ggaggccaag	gtggtacggt	cgcttgaagc	caaggagttc	521
	aagaccagcc	tggtcaacat	agcaaaacct	tgtctttaca	aaaagtaaaa	aaaaaaaaaa	581
	aaa						584
5							
	<210> 80						
	<211> 1351						
	<212> DNA						
	<213> Homo sapiens						
10							
	<220>						
	<221> CDS						
	<222> 259..831						
15							
	<220>						
	<221> sig_peptide						
	<222> 259..375						
	<223> Von Heijne matrix						
	score 5.809301698725						
20	seq FCVVCVIAIGVVQA/LI						
	<400> 80						
	aagctcccgg	ccgggctgac	tcaagcggag	gcgcgcggaa	cagtcgccga	ggcgattccc	60
	gcccagcagt	tcgacagaag	tgtacagagg	cttctggcaa	cacggattgc	cgtctacctg	120
25	atgacctttc	tcacgtgac	agtggcctgg	gcagcacaca	caaggttggt	ccaagttggt	180
	gggaaaacag	acgacacact	tgccctgctc	aacctggccg	catcatggct	gtgatgccct	240
	ccctccctcc	aggcctgc	atg atg acc atc acc ttc	ctg cct tac acg ttt			291
			Met Met Thr Ile Thr Phe Leu Pro Tyr Thr Phe				
			-35	-30			
30	tcg tta atg gtg acc ttc cct gat gtg cct ctg ggc atc ttc ttg ttc						339
	Ser Leu Met Val Thr Phe Pro Asp Val Pro Leu Gly Ile Phe Leu Phe						
		-25	-20	-15			
	tgt gtg tgt gtg atc gcc atc ggg gtc gtg cag gca ctg att gtg ggg						387
	Cys Val Cys Val Ile Ala Ile Gly Val Val Gln Ala Leu Ile Val Gly						
35		-10	-5	1			
	tac gca ttc cac ttc ccg cac ctg ctg agc ccg cag atc cag cgc tct						435
	Tyr Ala Phe His Phe Pro His Leu Leu Ser Pro Gln Ile Gln Arg Ser						
	5	10	15	20			
	gcc cac agg gct ctg tac cga cga cac gtc ctg ggc atc gtc ctc caa						483
40	Ala His Arg Ala Leu Tyr Arg Arg His Val Leu Gly Ile Val Leu Gln						
		25	30	35			
	ggc ccg gcc ctg tgc ttt gca gcg gcc atc ttc tct ctc ttc ttt gtc						531
	Gly Pro Ala Leu Cys Phe Ala Ala Ala Ile Phe Ser Leu Phe Phe Val						
		40	45	50			
45	ccc ttg tct tac ctg ctg atg gtg act gtc atc ctc ctc ccc tat gtc						579
	Pro Leu Ser Tyr Leu Leu Met Val Thr Val Ile Leu Leu Pro Tyr Val						
		55	60	65			
	agc aag gtc acc ggc tgg tgc aga gac agg ctc ctg ggc cac agg gag						627
	Ser Lys Val Thr Gly Trp Cys Arg Asp Arg Leu Leu Gly His Arg Glu						
50		70	75	80			
	ccc tcg gct cac cca gtg gaa gtc ttc tcg ttt gac ctc cac gag cca						675
	Pro Ser Ala His Pro Val Glu Val Phe Ser Phe Asp Leu His Glu Pro						
		85	90	95	100		
	ctc agc aag gag cgc gtg gaa gcc ttc agc gac gga gtc tac gcc atc						723
55	Leu Ser Lys Glu Arg Val Glu Ala Phe Ser Asp Gly Val Tyr Ala Ile						
		105	110	115			
	gtg gcc acg ctt ctc atc ctg gac atc tgc ccc tcc tgc tcc ctt tgg						771
	Val Ala Thr Leu Leu Ile Leu Asp Ile Cys Pro Ser Cys Ser Leu Trp						
		120	125	130			
60	ctg gct gtt gct tcc ttc cag cgt ctg ctc ctc cgc ggc ctc atc tgc						819
	Leu Ala Val Ala Ser Phe Gln Arg Leu Leu Leu Arg Gly Leu Ile Cys						
		135	140	145			
	ctc ttc gtc tgt tagagcgcgc gtctcgtctc agtcgtcacg tttttggttt						871

Leu Phe Val Cys
 150
 ttgtgggggt tttttttttt tttttttttg agacagtcct gctgtgtcgc ccaggctgga 931
 gtatagtggc tcaagctcag ctcaactgcaa cctccgcctc ccagggtcaa gcaattctcc 991
 5 tgcctcagcc tcccaagtag ttgggattac aagcaccac caccatgccc agctaacttt 1051
 ttgcattttt aatagagatg aggtttcacc aagttggcca ggctgggtctt gaactcctga 1111
 cctcaggtga tctgcccacc tcggcctccc aaagtgctgg gattacaggt gtaagccacc 1171
 gtgcccggcc atcgtaatgt ttgaatttgc ttttttacat cttccatcct tttggagtgt 1231
 cttgttcctt cgtcatagtt cagcactgtg accaccttgg ggtagacac tatggtttta 1291
 10 tatcctgtac ttgatattct cgagtccaag tctcctgatg ctctcaaaaa aaaaaaaaaa 1351

 <210> 81
 <211> 720
 <212> DNA
 15 <213> Homo sapiens

 <220>
 <221> CDS
 <222> 111..377
 20
 <220>
 <221> sig_peptide
 <222> 111..233
 <223> Von Heijne matrix
 25 score 5.26415334394122
 seq LWFLAQIPSRVAG/SL

 <400> 81
 aaaccgaaac cagcgctcca aacaattggg acccgggata ttatgccagt gaggctgtgc 60
 30 tgcggctgag cgggcctccc atccctctta aaagagttag gcatttagcc atg cct 116
 Met Pro
 -40
 ccc acc cgg gac cct ttc cag cag cct aca tta gat aac gat gat tcc 164
 Pro Thr Arg Asp Pro Phe Gln Gln Pro Thr Leu Asp Asn Asp Asp Ser
 35 -35 -30 -25
 tac tta gga gaa ctg cgg gct tcc aag gta ctg tgg ttt ctt gcg cag 212
 Tyr Leu Gly Glu Leu Arg Ala Ser Lys Val Leu Trp Phe Leu Ala Gln
 -20 -15 -10
 att ccc agt agg gtc gcc ggt agt ctt ctt tct gtc tgt gtg atg agc 260
 40 Ile Pro Ser Arg Val Ala Gly Ser Leu Leu Ser Val Cys Val Met Ser
 -5 1 5
 aga gat ggt aac ata aag gac tct ggt gaa gac act cag tcg ggt acc 308
 Arg Asp Gly Asn Ile Lys Asp Ser Gly Glu Asp Thr Gln Ser Gly Thr
 10 15 20 25
 45 agg gaa gtc tgt ttt ctg cct gcc tcc cta tct cca tat tca agt cgg 356
 Arg Glu Val Cys Phe Leu Pro Ala Ser Leu Ser Pro Tyr Ser Ser Arg
 30 35 40
 cta acg ttt cag agg cgt ttt tgagcagagg aaagtagagt tctagtctag 407
 Leu Thr Phe Gln Arg Arg Phe
 50 45
 aggaacaagg ggctctggca gctcaaatca attaaccaag atccaattcc ctggagaatt 467
 ttttaacctt cccactccac ccatcacttg cctgggtaac atcagacact ggatcaacct 527
 taaaaaggag tccatccaca gcatccaagg atccatagtg tcccctcaca ctgcagccac 587
 caatggaggc tactcccgaa agaaagatgg tggtctcttc tccacctagt gttgacagat 647
 55 ccttgaacta attatagtga aacatactgc ggcccacttc cattaaatag atttgtgcaa 707
 aaaaaaaaaa aaa 720

 <210> 82
 <211> 1029
 60 <212> DNA
 <213> Homo sapiens

 <220>

<221> CDS
 <222> 223..432

<220>
 5 <221> sig_peptide
 <222> 223..336
 <223> Von Heijne matrix
 score 4.17665217008018
 seq LVNVLFFFTPLMT/LV

10

<400> 82
 gtttttgtat tgggaagcagt tgtttggcct tgctgagcaa acgtctatgc cttctccatt 60
 acatccaaag gagaatagcc ccatgtgaag aatggaatca gtagatgttt ggtcgcgtgta 120
 ccatatccac tcctaggata caacaagagc aagcccaatt ctcttggtgg tgtgggcagt 180
 15 cggttgcac cacgtacctt tctcagctct ttttggaagc tt atg tcc tcc cca 234
 Met Ser Ser Pro
 -35
 caa ctt cca gct ttc tta tgg gac aag ggt aca ctc acc act gcc ata 282
 Gln Leu Pro Ala Phe Leu Trp Asp Lys Gly Thr Leu Thr Thr Ala Ile
 20 -30 -25 -20
 tct aat cct gct tgc ctg gta aat gtt ctc ttc ttc ttt aca ccc ctg 330
 Ser Asn Pro Ala Cys Leu Val Asn Val Leu Phe Phe Phe Thr Pro Leu
 -15 -10 -5
 atg act ctg gtc act cta ctc atc ctg gtc tgg aaa gta acc aaa gac 378
 25 Met Thr Leu Val Thr Leu Leu Ile Leu Val Trp Lys Val Thr Lys Asp
 1 5 10
 aaa agc aac aag aac aga gag aca cac cca aga aag gag gca aca tgg 426
 Lys Ser Asn Lys Asn Arg Glu Thr His Pro Arg Lys Glu Ala Thr Trp
 15 20 25 30
 30 ctg cca taaagatctg gatctcttgg tggggactcc actgaggtga agacctgatt 482
 Leu Pro
 gtacaagaga ggcacggcca ctggagctgt ctcagagccc agagccaggg gagccagagc 542
 tgcttttagcc accctgttcc tccattgcca gatgtccccc caggcctcat ttccttcctc 602
 tgccaccatc cctcttataa tgcactcctc ctgcggttct ttggcttgct ccagcttctg 662
 35 agtttgaatg tctttttttt tttttttttt tttttgkga tcttcaagac tgaaatagta 722
 aatggctctt gatttctgca ctaacagagg aaagaaacaa gtacatggaa aagtaaaaat 782
 tgattacaaa gcctaaattt tcctctataa attgggcatg tgctgactgt gggatattga 842
 aattattggg agctcacagc atctcaagtt atataatgaa gctattctgg aagctcattt 902
 ccagaagatc cttaaaatga aatggctcac tctctgctga attaatgttg agcaagttaa 962
 40 ctcctttttc aaatgaaatc caaattaaag aggcagtttt ttttgaaaaa ccaaaaaaaa 1022
 aaaaaaa 1029

<210> 83
 <211> 1788
 45 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 50 <222> 769..1272

<220>
 <221> sig_peptide
 <222> 769..843
 55 <223> Von Heijne matrix
 score 5.65786415517206
 seq AAHLLVVILPANA/AL

<400> 83
 60 ttgggcagaa aaattcaagc aggagattgt atttcttttg gagttgtacg atttccttta 60
 ttatttgaac tgcagtaaag aaagctggga tgggctcctc tagggatact tccagatccc 120
 tgggcggttg tagccctggc tcctctttaa atggatttgg tttcaaagac gatcatctcc 180
 gtcttctcgg atgtcatagt gccactgac atctccagct cctggccacc ctgggctttc 240

	tccactttgg	cctctatggt	ttgctttctc	accgtcttag	ccacgatata	tacctctctg	300
	tcatgtgatg	tgacccttgt	ttttgaacca	ggagtggccc	tgaggctcct	taaaaaagag	360
	ctgatcttac	tggttttctt	ttgtaaagct	cctgtgctag	atgcagattg	gttcttcccc	420
	agttcctgag	ttgttctcga	ctcctttttg	gtggagctgt	gggtggagct	cttgcgagag	480
5	gagccatgtc	gcttgcccct	tacgttgctg	tacccttttt	cttttttgtc	atctctcgtg	540
	tttttatggc	cagatgcgga	ccggtgggaa	gacgctttct	gattcttgtc	ccccgctctc	600
	ctgcggtgac	tttcacctgc	cttgcggtga	tgagaacttt	tcctactggg	atgtctgtcc	660
	ttcttttctc	ttcttttcct	gttttcatte	cagacttcag	cactgggctg	ggaaactttc	720
	tggcttccat	ctcgttcact	catgtagcct	tcgctttgca	aggtggag	atg agg ggt	777
10					Met Arg Gly		
					-25		
	ccc act gct ggt cct tca gtt ctt tct gct gca cac ttg ctg gtc gta	825					
	Pro Thr Ala Gly Pro Ser Val Leu Ser Ala Ala His Leu Leu Val Val						
	-20		-15		-10		
15	ata ctg cct gca aac gcc gca ctc aag ctg ctg tct tgg gag aga ctg	873					
	Ile Leu Pro Ala Asn Ala Ala Leu Lys Leu Leu Ser Trp Glu Arg Leu						
	-5		1		5		10
	gcg gcc ccc gcc atc gag gtg gaa gta cct tcc aag gag gtg ctt gca	921					
	Ala Ala Pro Ala Ile Glu Val Glu Val Pro Ser Lys Glu Val Leu Ala						
20		15		20		25	
	gca ccc acc aag gcc aag cta ata ccc tct gag gat atg ttg gca gca	969					
	Ala Pro Thr Lys Ala Lys Leu Ile Pro Ser Glu Asp Met Leu Ala Ala						
	30		35		40		
25	cct gcc atg gac ttg ctg gat tca ttt tct cct gga ttt ttg ata gct	1017					
	Pro Ala Met Asp Leu Leu Asp Ser Phe Ser Pro Gly Phe Leu Ile Ala						
	45		50		55		
	gct ccc gcc agc gct gtg atc act tgg cct ggg cct gca gat ttg gtt	1065					
	Ala Pro Ala Ser Ala Val Ile Thr Trp Pro Gly Pro Ala Asp Leu Val						
	60		65		70		
30	gtt gct atg ctc ata gca cct gtt gca gga ctc att gct gcc cct gct	1113					
	Val Ala Met Leu Ile Ala Pro Val Ala Gly Leu Ile Ala Ala Pro Ala						
	75		80		85		90
	att gcc aca tct gtt cta ggt cct gtt gct gtt cct gcc act gcc atg	1161					
	Ile Ala Thr Ser Val Leu Gly Pro Val Ala Val Pro Ala Thr Ala Met						
35		95		100		105	
	cca cct gct gtc ctt gct gct cct cct tca gca gcc cct gga gtg ctc	1209					
	Pro Pro Ala Val Leu Ala Ala Pro Pro Ser Ala Ala Pro Gly Val Leu						
	110		115		120		
40	gtg gat gga gaa gcc gca cta gcc gtt ccg tgg gag gca tgt tgg att	1257					
	Val Asp Gly Glu Ala Ala Leu Ala Val Pro Trp Glu Ala Cys Trp Ile						
	125		130		135		
	ccc tct ccc cca gca taagcagaag aggtggctgc agatacatca caaggcttgt	1312					
	Pro Ser Pro Pro Ala						
	140						
45	agagcccagt ctcactctga tccccttctc tgtggagctc tgcagcctat accaagggga	1372					
	agagaaacag atgagattga gatgactgaa agggagatca gaactttcta ctcctctctt	1432					
	atcctggagt taattcaagg gcttataatt agaagaacct gggtcggtgtg tgggtggctca	1492					
	cgcttgtaat cccaacactt tgggaggcca aggaggcgag atcgcttgag gccaggagtt	1552					
	caagaccagc cttgccaaaca tagcaaaacc ccgactctac taaaaataca aaaaattagc	1612					
50	tggacaggat ggcgcacatgcc tgtaatccca gctactcagt aggtctgaggt aggagtatcg	1672					
	cttgaactcg gatggcgagg gctgcagtga gccaaagactg cgccactcca ctgcactcca	1732					
	gcctgggcaa cagagtgaga cactgtttaa aaaaaagaaa gaaaaaaaaa aaaaaa	1788					
	<210> 84						
55	<211> 805						
	<212> DNA						
	<213> Homo sapiens						
	<220>						
60	<221> CDS						
	<222> 30..527						
	<220>						

<221> sig_peptide
 <222> 30..74
 <223> Von Heijne matrix
 score 8.68924532952647

5 seq PLLIICLLPAIEG/KN

<400> 84
 actggggcac agtaggagga acccagaag atg ctg cct ctc ctg atc atc tgt 53
 Met Leu Pro Leu Leu Ile Ile Cys
 -15 -10
 10 ctc ctg cct gcc att gaa ggg aag aac tgc ctc cgc tgc tgg cca gaa 101
 Leu Leu Pro Ala Ile Glu Gly Lys Asn Cys Leu Arg Cys Trp Pro Glu
 -5 1 5
 ctg tct gcc ttg ata gac tat gac ctg cag atc ctc tgg gtg acc cca 149
 15 Leu Ser Ala Leu Ile Asp Tyr Asp Leu Gln Ile Leu Trp Val Thr Pro 25
 10 15 20 25
 ggg cca ccc aca gaa ctt tct caa aat cgt gac cat ttg gaa gaa gaa 197
 Gly Pro Pro Thr Leu Ser Gln Asn Arg Asp His Leu Glu Glu Glu
 30 35 40
 20 aca gcc aaa ttc ttc act caa gta cac caa gcc att aaa acg tta cga 245
 Thr Ala Lys Phe Phe Thr Gln Val His Gln Ala Ile Lys Thr Leu Arg
 45 50 55
 gat gat aaa aca gta ctt ctg gaa gag atc tac acg cac aag aat ctc 293
 Asp Asp Lys Thr Val Leu Leu Glu Glu Ile Tyr Thr His Lys Asn Leu
 25 60 65 70
 ttt act gag agg ctg aat aag ata tct gat ggg ctg aag gag aag gac 341
 Phe Thr Glu Arg Leu Asn Lys Ile Ser Asp Gly Leu Lys Glu Lys Asp
 75 80 85
 ata cag tcc aca ctg aag gtc acc agc tgt gct gac tgc agg act cac 389
 30 Ile Gln Ser Thr Leu Lys Val Thr Ser Cys Ala Asp Cys Arg Thr His 100 105
 90 95 100 105
 ttc ctc tcc tgc aat gac ccc act ttc tgc cca gcc agg aac cgg cgg 437
 Phe Leu Ser Cys Asn Asp Pro Thr Phe Cys Pro Ala Arg Asn Arg Arg
 110 115 120
 35 acc tcc ctg tgg gct gtg agt ctc agc agt gct cta ctc ctg gcc ata 485
 Thr Ser Leu Trp Ala Val Ser Leu Ser Ser Ala Leu Leu Leu Ala Ile
 125 130 135
 gct gga gat gtt tct ttt act ggc aaa gga aga agg agg cag 527
 Ala Gly Asp Val Ser Phe Thr Gly Lys Gly Arg Arg Arg Gln
 40 140 145 150
 taaagcagga acagggcagc ccgcatgtct tccagaagtg aacagaggcc gcagctacca 587
 ccgtcacaaa gttcactcat ctctgggtcc cggtgacccc atccccccat accctccatc 647
 ctgggtcctg gggccccaaa gctctgaggc ctaggagact gcgctgtctc gtggtttgcc 707
 tactcctaca cctttgtaaa gagtctcttc attaaaaccc ctcttcataa aaaaaaaaaa 767
 45 aaaaaaaaaa aaaaaaaaaa aataaaaaaa aaaaaaaaaa 805

<210> 85
 <211> 814
 <212> DNA

50 <213> Homo sapiens

<220>
 <221> CDS
 <222> 39..506

55

<220>
 <221> sig_peptide
 <222> 39..83
 <223> Von Heijne matrix
 score 5.91494342964539
 60 seq ILMLTFIICGLLT/RV

<400> 85

	attcctcagg acacagagct tcctctctcc caggagcc atg aat atc ctg atg ctg	56
	Met Asn Ile Leu Met Leu	
	-15 -10	
5	acc ttc att atc tgt ggg ttg cta act cgg gtg acc aaa ggt agc ttt	104
	Thr Phe Ile Ile Cys Gly Leu Leu Thr Arg Val Thr Lys Gly Ser Phe	
	-5 1 5	
	gaa ccc caa aaa tgt tgg aag aat aat gta gga cat tgc aga aga cga	152
	Glu Pro Gln Lys Cys Trp Lys Asn Asn Val Gly His Cys Arg Arg Arg	
	10 15 20	
10	tgt tta gat act gaa agg tac ata ctt ctt tgt agg aac aag cta tca	200
	Cys Leu Asp Thr Glu Arg Tyr Ile Leu Leu Cys Arg Asn Lys Leu Ser	
	25 30 35	
	tgc tgc att tct ata ata tca cat gaa tat act cga cga cca gca ttt	248
	Cys Cys Ile Ser Ile Ile Ser His Glu Tyr Thr Arg Arg Pro Ala Phe	
15	40 45 50 55	
	cct gtg att cac cta gag gat ata aca ttg gat tat agt gat gtg gac	296
	Pro Val Ile His Leu Glu Asp Ile Thr Leu Asp Tyr Ser Asp Val Asp	
	60 65 70	
20	tct ttt act ggt tcc cca gta tct atg ttg aat gat ctg ata aca ttt	344
	Ser Phe Thr Gly Ser Pro Val Ser Met Leu Asn Asp Leu Ile Thr Phe	
	75 80 85	
	gac aca act aaa ttt gga gaa acc atg aca cct gag acc aat act cct	392
	Asp Thr Thr Lys Phe Gly Glu Thr Met Thr Pro Glu Thr Asn Thr Pro	
	90 95 100	
25	gag act act atg cca cca tcc gag gcc act act ccc gag act act atg	440
	Glu Thr Thr Met Pro Pro Ser Glu Ala Thr Thr Pro Glu Thr Thr Met	
	105 110 115	
	cca cca tct gag act gct act tcc gag act atg cca cca cct tct cag	488
	Pro Pro Ser Glu Thr Ala Thr Ser Glu Thr Met Pro Pro Pro Ser Gln	
30	120 125 130 135	
	aca gct ctt act cat aat taattaacat ttacttctgg tatggaacaa	536
	Thr Ala Leu Thr His Asn	
	140	
35	ctagaaatac tgctggaaat aatatccaaa gagctgatcc taccaatcca atttcaccag	596
	gaaaattcca tcagggattg gatgaccatg gggatggaca taattgctac taccaacaca	656
	acagccaaga gagttgcctt acaattagaa atgtgtagac agaaatgtat agaagataca	716
	aggattctct taattggact taaattcttt atctgtcttc ctccgatgta ctcaaatata	776
	tgagctaatt tttgtcttaa gtgaaaaaaaa aaaaaaaaa	814
40	<210> 86	
	<211> 598	
	<212> DNA	
	<213> Homo sapiens	
45	<220>	
	<221> CDS	
	<222> 115..429	
	<220>	
50	<221> sig_peptide	
	<222> 115..210	
	<223> Von Heijne matrix	
	score 8.2583062681354	
	seq LVAAMVLLSVVFC/LY	
55	<400> 86	
	attctaccag ctctggctga gctgagctt ccaaaagtga gctgagctgt tcaaccttgg	60
	atcttaatta ctcttagcag ggataattag gtccctcttt ctgagattac aggc atg	117
	Met	
60	gca aag atg ttt gat ctc agg acg aag atc atg atc ggc atc gaa agc	165
	Ala Lys Met Phe Asp Leu Arg Thr Lys Ile Met Ile Gly Ile Glu Ser	
	-30 -25 -20	
	agc tta ctg gtt gcc gcg atg gtg ctc cta agt gtt gtg ttc tgt ctt	213

```

Ser Leu Leu Val Ala Ala Met Val Leu Leu Ser Val Val Phe Cys Leu
-15          -10          -5          1
tac ttc aaa gta gct aag gca cta aaa gct gca aag gac cct gat gct      261
Tyr Phe Lys Val Ala Lys Ala Leu Lys Ala Ala Lys Asp Pro Asp Ala
5          5          10          15
gtg gct gta aaa aat cac aac cca gac aag gtg tgt tgg gcc acg aac      309
Val Ala Val Lys Asn His Asn Pro Asp Lys Val Cys Trp Ala Thr Asn
          20          25          30
agc cag gcc aaa gcc acc acc atg gag tct tgt cca tct ctc cag tgc      357
10 Ser Gln Ala Lys Ala Thr Thr Met Glu Ser Cys Pro Ser Leu Gln Cys
          35          40          45
tgt gaa ggt tgt aga atg cat gcc agt tct gat tcc ctg cca cct tgc      405
Cys Glu Gly Cys Arg Met His Ala Ser Ser Asp Ser Leu Pro Pro Cys
50          55          60          65
15 tgt tgt gac ata aat gag ggc ctc tgacttggga aagctgggca caaaaaatctt      459
Cys Cys Asp Ile Asn Glu Gly Leu
          70
catgagcaat atttctttct taatagaatg ttttattatt caagtcaagt tctagagtgt      519
ttacatacta ttatataatg tacagtgtta ttttctgtac ttctgaataa atgtgcaata      579
20 ttgcaaaaaa aaaaaaaaaa      598

<210> 87
<211> 699
<212> DNA
25 <213> Homo sapiens

<220>
<221> CDS
<222> 332..574
30
<220>
<221> sig_peptide
<222> 332..412
<223> Von Heijne matrix
35 score 7.96491294552426
seq ILGLFCCLPLAIP/AV

<400> 87
aatccctgt ggttggtgat caaggaagag catagtgcc aacctagggtg ccctcctggg      60
40 aatgttccag gagggcagga gtaggaggag gagtgttaga gtagagggga aatgatgaga      120
gcagaaagga gagtctcgct ctgtcaccca ggctggagtg cagtggcagg atcttggtc      180
acttcaacct ccacctcccg agttctgcct cagcctccca agtagctggg attacaggtc      240
cagtacttcc acgcttgag agtccaatta acaagagcaa gttctggtag aaagaagggtg      300
actttattcc agagctcagg tgtttgaact g atg tct gat gag gat gaa tcc      352
45 Met Ser Asp Glu Asp Glu Ser
          -25
agc gac tac ctc tgc ctg tcc atc ctg ggc ctc ttc tgt tgc ctt ccc      400
Ser Asp Tyr Leu Cys Leu Ser Ile Leu Gly Leu Phe Cys Cys Leu Pro
-20          -15          -10          -5
50 cta gcc atc cca gcc gtg atc ttt tct tgc ctg aca aag aac tac aat      448
Leu Ala Ile Pro Ala Val Ile Phe Ser Cys Leu Thr Lys Asn Tyr Asn
          1          5          10
aaa tcc agt gac tat gag ctg gca gcc aag acc tcc aaa caa gcc tac      496
Lys Ser Ser Asp Tyr Glu Leu Ala Ala Lys Thr Ser Lys Gln Ala Tyr
55          15          20          25
tac tgg gcc atc gcg agc atc act gtg gga atc tta ggt acc atc ttg      544
Tyr Trp Ala Ile Ala Ser Ile Thr Val Gly Ile Leu Gly Thr Ile Leu
          30          35          40
tac acc tac ctg ata tac tta ctt aga ttg taaactgctt cccagctctt      594
60 Tyr Thr Tyr Leu Ile Tyr Leu Leu Arg Leu
          45          50
gaacaaacca ccaaataat accacagtgc aattttaaaaa aaaaaaaaaa aaaaaaaaaa      654
aaaaaaaaa aaaaaaaaaa aaaaaaaaaa gaaaaaaaaa aaaaaa      699

```

<210> 88
 <211> 905
 <212> DNA
 5 <213> Homo sapiens

<220>
 <221> CDS
 <222> 133..417

10 <220>
 <221> sig_peptide
 <222> 133..213
 <223> Von Heijne matrix

15 score 11.106948594338
 seq LTSLILVTLISA/FV

<400> 88
 atttccaggg agctgaggag ctgagggcag agctagcttt tggttatttg ggatgttatt 60
 20 gccagtttcc tcccagggcc attgttacca cctgatcatt tgagtttttag tttctctagc 120
 agatgctgac ta atg act gac cag gat cga atc atc aat tta gtt gtt ggc 171
 Met Thr Asp Gln Asp Arg Ile Ile Asn Leu Val Val Gly
 -25 -20 -15
 agc tta aca tcc tta ttg att cta gta acg ctg ata agt gct ttt gtt 219
 25 Ser Leu Thr Ser Leu Leu Ile Leu Val Thr Leu Ile Ser Ala Phe Val
 -10 -5 1
 ttc cct caa cta cct cca aaa ccg ttg aat ata ttc ttt gct gtc tgc 267
 Phe Pro Gln Leu Pro Pro Lys Pro Leu Asn Ile Phe Phe Ala Val Cys
 5 10 15
 30 atc tct ttg agt agt att act gcc tgc ata atc tac tgg tat cga caa 315
 Ile Ser Leu Ser Ser Ile Thr Ala Cys Ile Ile Tyr Trp Tyr Arg Gln
 20 25 30
 gga gac tta gaa ccg aaa ttt aga aag cta att tac tat atc ata ttt 363
 Gly Asp Leu Glu Pro Lys Phe Arg Lys Leu Ile Tyr Tyr Ile Ile Phe
 35 35 40 45 50
 tct atc atc atg ttg tgt ata tgt gca aac ctg tac ttc cat gat gtg 411
 Ser Ile Ile Met Leu Cys Ile Cys Ala Asn Leu Tyr Phe His Asp Val
 55 60 65
 gga agg tgaggctgcc aaggagaagt acttaccagg actcttcaaa atgatacatt 467
 40 Gly Arg
 aggacagtga gtaatttttg gataaggat gctgaagaat ctctgcaga agtctgatac 527
 atgattttca tgttaattgt aaatgttaat tccctcttgc aaggagaca tatcctagat 587
 cactttgctt tttctttaag gagctgatgt tgcacctaaa cattccaacc cttaaagcta 647
 aaacagcaca aaaaaatttc acttttgaaa tgaaattttt ataattgtat ggcaaaaggc 707
 45 tatgtaaaaa caaatcttgc atcttaagac aaatattctt ttatttctgt taaactgaat 767
 atacaattgt tccctaggca accaactttt gcttataact acaatttaatt ttcacgttga 827
 caaaacacag tgaaaagaca actttgtgaa gatctaatta caataataaa taaaataatt 887
 tacaataaaaa aaaaaaaa 905

50 <210> 89
 <211> 514
 <212> DNA
 <213> Homo sapiens

55 <220>
 <221> CDS
 <222> 113..364

60 <220>
 <221> sig_peptide
 <222> 113..172
 <223> Von Heijne matrix
 score 4.37180298395146

seq SLLLSLPPHQGLT/FS

<400> 89
 ttttttacat ggtgttccca cagctgggag gacacccaca tggtcggcgt gcaggatatt 60
 5 tgcgtggacc ctagaaaagc caccacgacc tgtgggccat gatgctaccc ca atg gct 118
 Met Ala
 -20
 gct gct gct gtt cct tct ctt ctt ctt tct ctt cct cct cac cag ggg 166
 Ala Ala Ala Val Pro Ser Leu Leu Ser Leu Pro Pro His Gln Gly
 10 -15 -10 -5
 ctc act ttc tcc aac aaa ata caa cct ttt gga gct caa gga gtc ttg 214
 Leu Thr Phe Ser Asn Lys Ile Gln Pro Phe Gly Ala Gln Gly Val Leu
 1 5 10
 cat ccg gaa cca gga ctg cga gac tgg ctg ctg cca acg tgc tcc aga 262
 15 His Pro Glu Pro Gly Leu Arg Asp Trp Leu Leu Pro Thr Cys Ser Arg
 20 25 30
 caa ttg cga gtc gca ctg ccg gag aag ggg tcc gag ggc agt ctg tgt 310
 Gln Leu Arg Val Ala Leu Pro Glu Lys Gly Ser Glu Gly Ser Leu Cys
 35 40 45
 20 caa acg cag ctg cca gct act cca tgc ttc ctg cct tcg aat acg gtg 358
 Gln Thr Gln Leu Pro Ala Thr Pro Cys Phe Leu Pro Ser Asn Thr Val
 50 55 60
 aga acg tgaagtcag agctgctgct aaggcatgtg gcaaccttga agagaaggctc 414
 Arg Thr
 25 aagagctacc agccacaaaa agaatgccag cacttctgt gtctttgctt tggattcatg 474
 agaaatatac gttcctatatt gcttcaaaaa aaaaaaaaaa 514

<210> 90
 <211> 518
 30 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 35 <222> 9..380

<220>
 <221> sig_peptide
 <222> 9..104
 40 <223> Von Heijne matrix
 score 4.73369226787171
 seq AVFAVLFFVFFLFA/ML

<400> 90
 45 acatccta atg gtg gtg gtt gaa cca gga gcc agt tta ttc cca aat ggt 50
 Met Val Val Val Glu Pro Gly Ala Ser Leu Phe Pro Asn Gly
 -30 -25 -20
 gtt cct tgg ctc tat gct gtg ttt gct gtg ctt ttt gta ttt ttt ctt 98
 Val Pro Trp Leu Tyr Ala Val Phe Ala Val Leu Phe Val Phe Phe Leu
 50 -15 -10 -5
 ttt gcc atg tta tct ccc ttt tta ctt gag ata gac cag cac ata aag 146
 Phe Ala Met Leu Ser Pro Phe Leu Leu Glu Ile Asp Gln His Ile Lys
 1 5 10
 aaa ttc ttg atc aga tgc agg tat tct ctg cat aac act gtg cat aag 194
 55 Lys Phe Leu Ile Arg Cys Arg Tyr Ser Leu His Asn Thr Val His Lys
 15 20 25 30
 gac aaa aaa aac agt gag ata aag atg gac cat cta gaa agg cca ggc 242
 Asp Lys Lys Asn Ser Glu Ile Lys Met Asp His Leu Glu Arg Pro Gly
 35 40 45
 60 tgt cca ctg gag tca cca agg aga gga gtt ctg gga ggg aag aaa aat 290
 Cys Pro Leu Glu Ser Pro Arg Arg Gly Val Leu Gly Gly Lys Lys Asn
 50 55 60
 ggg atg gga aac gac cca tta cta ttt gtg aaa gtg aca aaa gaa ccc 338

Gly Met Gly Asn Asp Pro Leu Leu Phe Val Lys Val Thr Lys Glu Pro
65 70 75
agg gat tct gag gct gaa atc tat acc cct ggg cct tca gtt 380
Arg Asp Ser Glu Ala Glu Ile Tyr Thr Pro Gly Pro Ser Val
5 80 85 90
tgagagtcac ttagcctata tggaattacc tgtgacatta cattccagag agatgagaaa 440
ttctgagacc cttattatcg atgtttatat tgaaaaaatg gtaataaata ttttgagact 500
cccaaaaaaa aaaaaaaa 518

10 <210> 91
<211> 808
<212> DNA
<213> Homo sapiens

15 <220>
<221> CDS
<222> 155..340

<220>
20 <221> sig_peptide
<222> 155..292
<223> Von Heijne matrix
score 8.64329745298384
seq AVLLLLILFAIVFG/LL

25 <400> 91
cttttctct caacagttgc ttctttgagt caggggtgcag ctctgggtcac ctggcggcct 60
cttcagctca gccctccaca aagtgtgagc ctgaaggacc accctgaatt gccctttag 120
gaccagaac agctaccagc agaatcagat tctc atg gac caa ctg gta ttc aaa 175
30 Met Asp Gln Leu Val Phe Lys
-45 -40
gag aca atc tgg aat gat gcg ttc tgg cag aac ccc tgg gac cag ggg 223
Glu Thr Ile Trp Asn Asp Ala Phe Trp Gln Asn Pro Trp Asp Gln Gly
-35 -30 -25
35 ggc ctg gca gtg att atc tta ttc atc acc gct gtc ctg ctt ctc atc 271
Gly Leu Ala Val Ile Ile Leu Phe Ile Thr Ala Val Leu Leu Leu Ile
-20 -15 -10
tta ttt gcc atc gtg ttt ggt tta ctc act tcc aca gaa aac act cag 319
Leu Phe Ala Ile Val Phe Gly Leu Leu Thr Ser Thr Glu Asn Thr Gln
40 -5 1 5
tgt gaa gcg ggt gaa gag gag tgacctgact tgctggggac tgagatggca 370
Cys Glu Ala Gly Glu Glu
10 15
gcaggggagg cgagctgacc tgccccatt ccagtggtgg gccccttcgc ggttcctct 430
45 ggctcagggg ccaagccctg gtgtcttctt tcccaccag gaaaaagtct agtaaaatac 490
tgtatctggc ttagggttg tcagactagt aagatgggga ggctgggtctg agaccaattc 550
tggtccttg accctattgt ttttagggtt ccccgaccag aaccctaaaa gcacatggag 610
aggatggctc cactgcctca ggtggaagga gctatggcta acaagggttct ctaacaggct 670
cacaggccca gccagcaatt tcacaaatcc ttgacagaga aagacacaac caaatgaaat 730
50 aaaaattcct tttcaaatct gctaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 790
aaaaaaaaaa aaaaaaaaaa 808

<210> 92
<211> 737
55 <212> DNA
<213> Homo sapiens

<220>
<221> CDS
60 <222> 185..634

<220>
<221> sig_peptide

<222> 185..253
 <223> Von Heijne matrix
 score 9.49395175807817
 seq SLLFICFFGESFC/IC

5

<400> 92
 atatttttgct gactggcaag gttatatgaa gtgctttttat tgaagcacca ttttaactaa 60
 tagctcctgg tatttttctgc ttccttcgt aggggaattta gttattttat tttattattt 120
 agctaattta gctatttttaa aatagctaaa ttttagctac ttttttttca attgacaaag 180
 10 aagg atg tct aat caa aga cta ccg ctg att ttt tct ctg ttg ttt atc 229
 Met Ser Asn Gln Arg Leu Pro Leu Ile Phe Ser Leu Leu Phe Ile
 -20 -15 -10
 tgc ttc ttc ggg gag agt ttc tgc att tgt gat gga act gtc tgg aca 277
 Cys Phe Phe Gly Glu Ser Phe Cys Ile Cys Asp Gly Thr Val Trp Thr
 -5 1 5
 15 aag gtt gga tgg gag att ctt cca gaa gaa gta cat tat tgg aaa ggt 325
 Lys Val Gly Trp Glu Ile Leu Pro Glu Glu Val His Tyr Trp Lys Gly
 10 15 20
 tgt tta tat ctc att tat aat tta tta caa gct gtc ttc ttc gtc tta 373
 20 Cys Leu Tyr Leu Ile Tyr Asn Leu Leu Gln Ala Val Phe Phe Val Leu
 25 30 35 40
 ttt gtt ttg tct gtg cat tac ctg tgg aag aaa tgg aag aaa cac caa 421
 Phe Val Leu Ser Val His Tyr Leu Trp Lys Lys Trp Lys Lys His Gln
 45 50 55
 25 aaa aag ctg aaa aag caa gcc tcc tta gaa aaa cct ggt aat gat cta 469
 Lys Lys Leu Lys Lys Gln Ala Ser Leu Glu Lys Pro Gly Asn Asp Leu
 60 65 70
 gaa agc cca ttg atc aac aac att gac caa aca ctc cac aga gtg gca 517
 Glu Ser Pro Leu Ile Asn Asn Ile Asp Gln Thr Leu His Arg Val Ala
 75 80 85
 30 acc aca gca tca gtg ata tac aag atc tgg gag cac agg tct cac cat 565
 Thr Thr Ala Ser Val Ile Tyr Lys Ile Trp Glu His Arg Ser His His
 90 95 100
 cct tcc tct aag aaa att aag cac tgc aaa tta aag aag aag agt aaa 613
 35 Pro Ser Ser Lys Lys Ile Lys His Cys Lys Leu Lys Lys Lys Ser Lys
 105 110 115 120
 gaa gaa gga gcc aga aga tac taaataaatg catatgcaaa tgtagcttag 664
 Glu Glu Gly Ala Arg Tyr
 125
 40 tcaattatag atatcacaaa agaaatctat catctaagga ttaaaaattg ttctttggaa 724
 aaaaaaaaaa aaa 737

<210> 93
 <211> 728
 45 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 50 <222> 53..646

<220>
 <221> sig_peptide
 <222> 53..91
 55 <223> Von Heijne matrix
 score 4.95353272042967
 seq MLLGRLTSQLLRA/VP

<400> 93
 60 aatttgagcc gcgtcgagct cccttgggac ctgtggccgc cgccacaga cc atg ctc 58
 Met Leu
 ctg ggg cgc ctg act tcc cag ctg ttg agg gcc gtt cct tgg gca ggc 106
 Leu Gly Arg Leu Thr Ser Gln Leu Leu Arg Ala Val Pro Trp Ala Gly

	-10		-5		1		5		
	ggc cgc ccg cct tgg ccc gtc tct gga gtg ctg ggc agc cgg gtc tgc	154							
	Gly Arg Pro Pro Trp Pro Val Ser Gly Val Leu Gly Ser Arg Val Cys								
		10		15			20		
5	ggg ccc ctt tac agc aca tcg ccg gcc ggc cca ggt agg gcg gcc tct	202							
	Gly Pro Leu Tyr Ser Thr Ser Pro Ala Gly Pro Gly Arg Ala Ala Ser								
		25		30			35		
	ctc cct cgc aag ggg gcc cag ctg gag ctg gag gag atg gtc ccc agg	250							
	Leu Pro Arg Lys Gly Ala Gln Leu Glu Leu Glu Glu Met Val Pro Arg								
10		40		45			50		
	aag atg tcc gtc agc ccc ctg gag agc tgg ctc acg gcc cgc tgc ttc	298							
	Lys Met Ser Val Ser Pro Leu Glu Ser Trp Leu Thr Ala Arg Cys Phe								
		55		60			65		
	ctg ccc aga ctg gat acc ggg acc gca ggg act gtg gct cca ccg caa	346							
15	Leu Pro Arg Leu Asp Thr Gly Thr Ala Gly Thr Val Ala Pro Pro Gln								
		70		75			80		85
	tcc tac cag tgt ccg ccc agc cag ata ggg gaa ggg gcc gag cag ggg	394							
	Ser Tyr Gln Cys Pro Pro Ser Gln Ile Gly Glu Gly Ala Glu Gln Gly								
		90		95			100		
20	gat gaa ggc gtc gcg gat gcg cct caa att cag tgc aaa aac gtg ctg	442							
	Asp Glu Gly Val Ala Asp Ala Pro Gln Ile Gln Cys Lys Asn Val Leu								
		105		110			115		
	aag atc cgc cgg cgg aag atg aac cac cac aag tac cgg aag ctg gtg	490							
	Lys Ile Arg Arg Arg Lys Met Asn His His Lys Tyr Arg Lys Leu Val								
25		120		125			130		
	aag aag acg cgg ttc ctg cgg agg aag gtc cag gag gga cgc ctg aga	538							
	Lys Lys Thr Arg Phe Leu Arg Arg Lys Val Gln Glu Gly Arg Leu Arg								
		135		140			145		
	cgc aag cag atc aag ttc gag aaa gac ctg agg cgc atc tgg ctg aag	586							
30	Arg Lys Gln Ile Lys Phe Glu Lys Asp Leu Arg Arg Ile Trp Leu Lys								
		150		155			160		165
	gcg ggg cta aag gaa gcc ccc gaa ggc tgg cag acc ccc aag atc tac	634							
	Ala Gly Leu Lys Glu Ala Pro Glu Gly Trp Gln Thr Pro Lys Ile Tyr								
		170		175			180		
35	ctg cgg ggc aaa tgagtctggc gccgcccttc ccgcccggtt ctgctgtgat	686							
	Leu Arg Gly Lys								
		185							
	ccgtagtaat aaattctcag aggacccaaa aaaaaaaaaa aa	728							
40	<210> 94								
	<211> 582								
	<212> DNA								
	<213> Homo sapiens								
45	<220>								
	<221> CDS								
	<222> 247..510								
	<220>								
50	<221> sig_peptide								
	<222> 247..318								
	<223> Von Heijne matrix								
	score 5.20026065148038								
	seq FCALEVVLPSDC/RS								
55	<400> 94								
	atcataactca ccattggctca caaactgcct gtttgaaaact cccttcagtt ctgagaggat	60							
	gggaacattc ttttaagcggc tcgtcttggc acgagacata aggcaattca acatcaagcc	120							
	cttgccctga acagttccaa atgccaagaa ctggcgaatt actactttgg tttcaatggg	180							
60	tgttccaaaa ggatcatcaa gcttcaggag ctttctgacc ttgaagaaag ggaaaatgaa	240							
	gatagc atg gtg cca ctt ccg aag caa agc ctg aag ttc ttc tgt gct	288							
	Met Val Pro Leu Pro Lys Gln Ser Leu Lys Phe Phe Cys Ala								
		-20		-15					

```

tta gaa gtg gtg ttg cca tcc tgt gat tgc agg agt cct ggc att ggc      336
Leu Glu Val Val Leu Pro Ser Cys Asp Cys Arg Ser Pro Gly Ile Gly
-10          -5          1          5
ttg gtg gag gag cct atg gat aag gtg gag gaa gga cca tta tca ttc      384
5 Leu Val Glu Glu Pro Met Asp Lys Val Glu Glu Gly Pro Leu Ser Phe
10          10          15          20
ctt atg aaa agg aag aca gcc cag aag ctt gct att cag aag gct ttg      432
Leu Met Lys Arg Lys Thr Ala Gln Lys Leu Ala Ile Gln Lys Ala Leu
25          30          35
10 tca gat gca ttc cag aaa ctg ttg att gtt gtt cta ggt aag act gtc      480
Ser Asp Ala Phe Gln Lys Leu Leu Ile Val Val Leu Gly Lys Thr Val
40          45          50
ttg atc atc ctt gaa gta ctt cag ttt cag taagcaaata aactcatttt      530
Leu Ile Ile Leu Glu Val Leu Gln Phe Gln
15 55          60
gaaaagttaa ttgaataaaa atattgatat ctaaagcaaa aaaaaaaaaa aa      582

<210> 95
<211> 1913
20 <212> DNA
    <213> Homo sapiens

<220>
<221> CDS
25 <222> 143..592

<220>
<221> sig_peptide
<222> 143..277
30 <223> Von Heijne matrix
    score 5.94057630118762
    seq VLVDLAILGQAYA/FA

<400> 95
35 attttttttgt gcctaagatg cccagtgcgt tgctggggttt ttctgctgtc ctcgggctct      60
   ggacatgagg ccagaccttg tgaccttggt ggcagtgggc agtggttga tgtgaggtcc      120
   cagagacggc aggttcatca ag atg gtg ctc atg tgg acc agt ggt gac gcc      172
                        Met Val Leu Met Trp Thr Ser Gly Asp Ala
                        -45          -40
40 ttc aag acg gcc tac ttc ctg ctg aag ggt gcc cct ctg cag ttc tcc      220
   Phe Lys Thr Ala Tyr Phe Leu Leu Lys Gly Ala Pro Leu Gln Phe Ser
   -35          -30          -25          -20
   gtg tgc ggc ctg ctg cag gtg ctg gtg gac ctg gcc atc ctg ggg cag      268
   Val Cys Gly Leu Leu Gln Val Leu Val Asp Leu Ala Ile Leu Gly Gln
45          -15          -10          -5
   gcc tac gcc ttc gcc cca ccc cca gaa gcc ggc gcc cca cgc cgt gca      316
   Ala Tyr Ala Phe Ala Pro Pro Pro Glu Ala Gly Ala Pro Arg Arg Ala
                        1          5          10
50 ccc cac tgg cac caa ggc cct ctg aca gtg ggg agg acg agg atg tgg      364
   Pro His Trp His Gln Gly Pro Leu Thr Val Gly Arg Thr Arg Met Trp
   15          20          25
   gac cgc cag ccg cgg gca ctg gtg ggc cct gac ctc ccc gcg ggg agg      412
   Asp Arg Gln Pro Arg Ala Leu Val Gly Pro Asp Leu Pro Ala Gly Arg
   30          35          40          45
55 gtg ggt gcc gtg gcc cct gca ggt gtg gca gag atg ggg cac ggg cat      460
   Val Gly Ala Val Ala Pro Ala Gly Val Ala Glu Met Gly His Gly His
                        50          55          60
   tgg ggt ctc cat cag cct ctg tgg ggt gtc tca ggg tgg gca gtg ggg      508
   Trp Gly Leu His Gln Pro Leu Trp Gly Val Ser Gly Trp Ala Val Gly
60          65          70          75
   gtg ggg ctg gga cgc tgt ttg tgc tca gcg ggg aca gcc agg gtt gat      556
   Val Gly Leu Gly Arg Cys Leu Cys Ser Ala Gly Thr Ala Arg Val Asp
                        80          85          90

```

	ctg gcc ccg agg gtt ttg gat gtt ttt agg atg	aca taaaaagcaa	602
	Leu Ala Pro Arg Val Leu Asp Val Phe Arg Met	Thr	
	95	100	105
	gtgtttttccc catttcctct tatgaaacac cgtctgagcc	caaggtacac attgggcggc	662
5	ctgcaggaac ctgctccagg tggacacacg ggccagcagc	cgcgaacctt gaagctgggg	722
	tgaccgcagg agaccctgta aggcctgtga gcggagccct	cgaccccgctg acaccctggc	782
	cagacaccct gcttggactg ggggtggcctc tgctaccag	gggtctggca cgggggagg	842
	ctggggcttt ctctgcctgg tacacacgga aaggcggctg	tgcgacgca gggtcaccgt	902
	gctccgggtt ttctgacagt cgggtgtttcc tgggcctttg	gagtggctgc gaggcctgaa	962
10	cgccttgtgg atccgctgtg tccagcccgg ctgagcatcg	ccagggctag ctcatgctgc	1022
	tcttgtcagc ctctggttct cctcgagtcc ttggggacgt	ggcagatgcc agcgaccatc	1082
	agacaacgtg gaggccctca tgggcaatgg ctgagggggc	cgggctgagg ctgtgcacat	1142
	gcagtctgca cgccactctt gggctctgct ggcgagatc	cccttccttc tgggtgcaga	1202
	ctgcacctcc ggatgcagtt ttgatgtcca tcttccagga	gagagacggc ctcgggtcca	1262
15	gggagtggag ggggctgccc ctgccgtgca ggtcctggcc	gatggcgcct taccctgctg	1322
	ccctgggctt ttggcctgaa gcaaattcct gagtgggggg	tactggggcc tgcgcacatc	1382
	tgtcctgtcc actgcccacc cccgtgtgct ggctccctca	cttctggctg cagtgggagc	1442
	cgccagtctg acccttgtca ccgcacgctc tgcccccacc	ccgttgcaag aggtcacacc	1502
	atgtcagcag ccttgcactg accgcagccg gccccaggc	ctcagagttc tggatgcttc	1562
20	cgtgcggctc caacaggcat cgtcttccct tccgcaggtg	gaggggcccgc ttcccgcagg	1622
	catctgagct ctgtgccggg gccgtggcca tgggaagatg	ttccacgctg cctcctcctc	1682
	gagttttcct cggaacactc cttgaatgtc tgagtggagg	tcctgcttag ctctttggcc	1742
	tgtgagatgc tttgaaaatt tttatttttt taagatgaag	caagatgtct gtagcggtaa	1802
	ttgcctcaca ttaaactgtc gccgactgca ggcgagtg	ctgctgaatg taccctgtgt	1862
25	ggcgacttgg aatcaataaa ccatttgtgg atcctaaaaa	aaaaaaaaa a	1913
	<210> 96		
	<211> 670		
	<212> DNA		
30	<213> Homo sapiens		
	<220>		
	<221> CDS		
	<222> 33..458		
35	<220>		
	<221> sig_peptide		
	<222> 33..89		
	<223> Von Heijne matrix		
40	score 6.45239823575329		
	seq SVFLLMVNGQVES/AQ		
	<400> 96		
45	aggtgggtcc ccccggcacc cccagacctg cc atg gcg acc gcg agt cct agc	53	
	Met Ala Thr Ala Ser Pro Ser		
	-15		
	gtc ttt cta ctc atg gtc aac ggg cag gtg gag agc gcc cag ttt cca	101	
	Val Phe Leu Leu Met Val Asn Gly Gln Val Glu Ser Ala Gln Phe Pro		
	-10 -5 1		
50	gag tat gat gac ttc tac tgc aag tac tgc ttt gtg tac ggc cag gac	149	
	Glu Tyr Asp Asp Phe Tyr Cys Lys Tyr Cys Phe Val Tyr Gly Gln Asp		
	5 10 15 20		
	tgg gcc ccc aca gcg ggt ctg gag gag ggg atc tca cag atc aca tcc	197	
	Trp Ala Pro Thr Ala Gly Leu Glu Glu Gly Ile Ser Gln Ile Thr Ser		
55	25 30 35		
	aag agc caa gat gtg cgg caa gca ctg gtg tgg aac ttc ccc att gat	245	
	Lys Ser Gln Asp Val Arg Gln Ala Leu Val Trp Asn Phe Pro Ile Asp		
	40 45 50		
	gtc acc ttt aaa agc acc aac ccc tac ggc tgg cca cag atc gtg ctc	293	
60	Val Thr Phe Lys Ser Thr Asn Pro Tyr Gly Trp Pro Gln Ile Val Leu		
	55 60 65		
	agc gtg tat gga cca gat gtg ttc ggg aac gat gtg gtt cga ggc tat	341	
	Ser Val Tyr Gly Pro Asp Val Phe Gly Asn Asp Val Val Arg Gly Tyr		

	70	75	80	
	ggg gcc gtg cac gtg ccc ttc tca cct ggc cgg cac aaa agg acc atc			389
	Gly Ala Val His Val Pro Phe Ser Pro Gly Arg His Lys Arg Thr Ile			
	85	90	95	100
5	ccc atg ttt gtc cca gaa tct acg tct aaa ctg cag aag ttt aca aga			437
	Pro Met Phe Val Pro Glu Ser Thr Ser Lys Leu Gln Lys Phe Thr Arg			
		105	110	115
	tct gca agc tgc tcc acc cac tgaggacaaa tagaaacagg tcccctggga			488
	Ser Ala Ser Cys Ser Thr His			
10		120		
	gtgctgagtc acggggctcc cttcagccct gttccagcag cagaaggccg ggcgatttta			548
	ccctgtgccc tgtgaaaaat ctttgtgtct gagggggcag aggaaaaact cttgtcagat			608
	gggaaaaatg ctcattgacat aatgtgacat taaaagggtgg gaaacaaaaa aaaaaaaaaa			668
	aa			670
15	<210> 97			
	<211> 939			
	<212> DNA			
	<213> Homo sapiens			
20	<220>			
	<221> CDS			
	<222> 1..336			
25	<220>			
	<221> sig_peptide			
	<222> 1..81			
	<223> Von Heijne matrix			
	score 3.68137078794859			
30	seq AHLCSDSLPEQQ/QD			
	<400> 97			
	act tcc gaa gag aga acc gcc atg aag aga gaa ggg ggt gcc gcc cac			48
	Thr Ser Glu Glu Arg Thr Ala Met Lys Arg Glu Gly Ala Ala His			
35		-25	-20	-15
	ctc tgc tcc gac agc ctc ccg gag tcc cag cag caa gac ggc aac cac			96
	Leu Cys Ser Asp Ser Leu Pro Glu Ser Gln Gln Gln Asp Gly Asn His			
		-10	-5	1
	gca ccc aac ttc tcc agc cac ggc tca tgc cgc cgt cgc cag cgg scc			144
40	Ala Pro Asn Phe Ser Ser His Gly Ser Cys Arg Arg Arg Gln Arg Xaa			
		10	15	20
	gac atg aca agg cgc tgc atg ccc gct agg cca ggt ttc ccc tca tcc			192
	Asp Met Thr Arg Arg Cys Met Pro Ala Arg Pro Gly Phe Pro Ser Ser			
		25	30	35
45	cca gcc ccg ggg tcg tcg ccc ccg cgc tgc cat ctg aga ccc ggt agt			240
	Pro Ala Pro Gly Ser Ser Pro Pro Arg Cys His Leu Arg Pro Gly Ser			
		40	45	50
	acc gcc cat gct gca gcg gga aag aga aca gag agt cct ggg gac agg			288
	Thr Ala His Ala Ala Ala Gly Lys Arg Thr Glu Ser Pro Gly Asp Arg			
50		55	60	65
	tac cgt gca gag ggc ttg aga agg ggc cgg gtc gcg ggg gca agg gta			336
	Tyr Arg Ala Glu Gly Leu Arg Arg Gly Arg Val Ala Gly Ala Arg Val			
		70	75	80
	tgaggggagg gctgcagacc gccgctcttc cagttccgc catcctccgc gagctcaggc			396
55	gttggcattt cggggcctgg caaatccccg cccgcctcc gcgcaggggc tactgggagt			456
	tggagtttgc ttctctgtag ttgggcagct gctcttggtc tagtgaccac cagcctggac			516
	agctacggag aaccgcctt aggtagaaag aaagtgattt ttttcctttg caagagtttg			576
	accggggacc ctaactgctt aatgcatatt tagatcgttt tctgtacgtt gtcagttcta			636
	ctgatccctag tggtttagta atataaacct tttctatggt gtgggtgaaa ttatgtaacc			696
60	tgtgatgagg gaatcccttc cacgaattac tttgtagtcc agcgtgcacg ctagtccata			756
	cttaaaagaa cttgcagatt tggaatgtga cgtgttttct ctttcagtaa cttcacgcct			816
	ctccaagagg ctaatttttt tgtaaagatt ttgtgggagc tatgtaatga gatggggagt			876
	ttcatctaatt gacatcctct gacaataaaa aatgttttaaa ttccccaaaa aaaaaaaaaa			936

aaa

939

<210> 98
 <211> 661
 5 <212> DNA
 <213> Homo sapiens

 <220>
 <221> CDS
 10 <222> 174..443

 <220>
 <221> sig_peptide
 <222> 174..269
 15 <223> Von Heijne matrix
 score 4.13107367257584
 seq SSLAFCQVGFLTA/QP

 <400> 98
 20 aaaaaggaac tttcagtgat aatgaacaaa actcaggagc tatgtggatg acaggagcac 60
 ctagatgacc gactttaccc acttcaaagt ctaccttgac cctagcactc tctccaccct 120
 gcatcctcac ctgagaccat cagttgggta ggccaacagc tcaccatcaa ttc atg 176
 Met
 ccc tgc cta gac caa cag ctc act gtt cat gcc cta ccc tgc cct gcc 224
 25 Pro Cys Leu Asp Gln Gln Leu Thr Val His Ala Leu Pro Cys Pro Ala
 -30 -25 -20
 cag ccc tcc tct ctg gcc ttc tgc caa gtg ggg ttc tta aca gca cag 272
 Gln Pro Ser Ser Leu Ala Phe Cys Gln Val Gly Phe Leu Thr Ala Gln
 -15 -10 -5 1
 30 cct tca cct ccg aga agg cgc aat ggg aaa gac aga tac acg ttg gtt 320
 Pro Ser Pro Pro Arg Arg Arg Asn Gly Lys Asp Arg Tyr Thr Leu Val
 5 10 15
 ctg caa cac cag gaa tgc cag gat gat tta gcc acc tcc tca ctt gtc 368
 Leu Gln His Gln Glu Cys Gln Asp Asp Leu Ala Thr Ser Ser Leu Val
 35 20 25 30
 tac ctt tcc ctc ccc tgc ttc aaa gac ttg ggt cga tcg aag cac caa 416
 Tyr Leu Ser Leu Pro Cys Phe Lys Asp Leu Gly Arg Ser Lys His Gln
 35 40 45
 agc atc act gtt gct gac act aac aag tagtgccaag ggattgcctt 463
 40 Ser Ile Thr Val Ala Asp Thr Asn Lys
 50 55
 taaggaagat caggagcgga acatctggtg gcaaagaaaa tctttctaata agccccattc 523
 tagtgaccac cttcaacctc ctcatagcag gagagtttgg gagtagggga cttaggatgt 583
 tttgttcttt taatcaattc agaaaatatg tatgtttgaa ataaaaataa aaataactga 643
 45 gccaaaaaaa aaaaaaaa 661

 <210> 99
 <211> 647
 <212> DNA
 50 <213> Homo sapiens

 <220>
 <221> CDS
 <222> 282..521
 55
 <220>
 <221> sig_peptide
 <222> 282..386
 <223> Von Heijne matrix
 60 score 3.64439944832387
 seq LEPGLSSSAACNG/KE

 <400> 99

	acttgcgtgt caccgttacc gtagegactg ggcttctgga ctgtatatcc tagctgcctt	60
	gtcaacatct tcgagcatcg gcagctccgg aggccggggt aactggcagg taggaaacta	120
	tgtgaaagaa tctcctgatg tcataatttc cgggtgtcac cggaacattt gatcatcatt	180
	cctttggcaa ttccagcctt ctgtggaaag gccagtagaa agcattgatt tattcacctc	240
5	tacaggaatc agactcagcc tcttttggtt ttcagtgaag t atg cct ttt caa ttt	296
	Met Pro Phe Gln Phe	
	-35	
	gga acc cag cca agg agg ttt cca gtg gaa gga gga gat tct tca att	344
	Gly Thr Gln Pro Arg Arg Phe Pro Val Glu Gly Gly Asp Ser Ser Ile	
10	-30 -25 -20 -15	
	gag ctg gaa cct ggg ctg agc tcc agt gct gcc tgt aat ggg aag gag	392
	Glu Leu Glu Pro Gly Leu Ser Ser Ser Ala Ala Cys Asn Gly Lys Glu	
	-10 -5 1	
	atg tca cca acc agg caa ctc cgg agg tgc cct gga agt cat tgc ctg	440
15	Met Ser Pro Thr Arg Gln Leu Arg Arg Cys Pro Gly Ser His Cys Leu	
	5 10 15	
	aca ata act gat gtt ccc gtc act gtt tat gca aca acg aga aag cca	488
	Thr Ile Thr Asp Val Pro Val Thr Val Tyr Ala Thr Thr Arg Lys Pro	
	20 25 30	
20	cct gca caa agc agc aag gaa atg cat cct aaa tagcaccatt aagtcttttg	541
	Pro Ala Gln Ser Ser Lys Glu Met His Pro Lys	
	35 40 45	
	tcaaggtctg actaggtcaa gggtaatgga ccagtatcat ctgggtgatct ggtaaacaaa	601
	taaaagtggg ggcaccttta gatgatgaca aaaaaaaaaa aaaaaa	647
25		
	<210> 100	
	<211> 1006	
	<212> DNA	
	<213> Homo sapiens	
30		
	<220>	
	<221> CDS	
	<222> 251..643	
35		
	<220>	
	<221> sig_peptide	
	<222> 251..295	
	<223> Von Heijne matrix	
	score 3.74215118492367	
40	seq LLMFTQLLLCGFL/YV	
	<400> 100	
	aggaagccag agggctggaa atacagcagc ctttgaagta ccctctgtta atttgatgg	60
	atctcagtgt gccccgttcg agacctctcc accaacacct tctgatcttg cgatttgctc	120
45	ttcttgactt taattagtat ctaggaaagt ctaaactttg gacctacctc tttttttgat	180
	actcattttt gtacttttgc tctctgggat tggtttctta aagaatctgg atccttttta	240
	atatgtcaaa atg agt ctg ctg atg ttt aca caa cta ctg ctc tgt gga	289
	Met Ser Leu Leu Met Phe Thr Gln Leu Leu Leu Cys Gly	
	-15 -10 -5	
50	ttt tta tat gtt cgg gtt gat gga tcg cgt ctt cgc cag gag gac ttt	337
	Phe Leu Tyr Val Arg Val Asp Gly Ser Arg Leu Arg Gln Glu Asp Phe	
	1 5 10	
	ccc ccg cgg att gtg gag cat cct tcc gat gtc atc gtc tct aag ggc	385
	Pro Pro Arg Ile Val Glu His Pro Ser Asp Val Ile Val Ser Lys Gly	
55	15 20 25 30	
	gag ccc acg act ctg aac tgc aag gcg gag ggc cgg cca acg ccc acc	433
	Glu Pro Thr Thr Leu Asn Cys Lys Ala Glu Gly Arg Pro Thr Pro Thr	
	35 40 45	
	att gag tgg tac aaa gat ggg gag cga gtg gag act gac aag gac gat	481
60	Ile Glu Trp Tyr Lys Asp Gly Glu Arg Val Glu Thr Asp Lys Asp Asp	
	50 55 60	
	ccc cgg tcc cac agg atg ctt ctg ccc agc gga tcc tta ttc ttc ttg	529
	Pro Arg Ser His Arg Met Leu Leu Pro Ser Gly Ser Leu Phe Phe Leu	


```

      65              70              75
cgc atc gtg cac ggg cgc agg agt aaa cct gat gaa gga agc tac gtt      577
Arg Ile Val His Gly Arg Arg Ser Lys Pro Asp Glu Gly Ser Tyr Val
      80              85              90
5  tgt gtt gcg agg aac tat ctt ggt gaa gca gtg agt cga aat gcg tct      625
Cys Val Ala Arg Asn Tyr Leu Gly Glu Ala Val Ser Arg Asn Ala Ser
      95              100              105              110
ctg gaa gtg gca tgt aag tgaacataat gaacctcatg tgcacattta      673
Leu Glu Val Ala Cys Lys
10              115
ctttttattta tttcaagtaa gttttgatgt gttcccatag acgctgaaac ctaaagaatc      733
aatcaacaca ctgcataatt ttacttggtc ttcttcagag aagtctgggc aagatagtat      793
caagccaggg tgttgtagta agtttggtta tatgaaatca agatgaccaa tatgttatta      853
taagaaagca ggccggggcgc ggtgggtcac gcctgtaatc ccagcacttt gggaggcgga      913
15 ggccggcgga tcacgaggtc aggagatcga gaccatcctg ggtagcacgg tggggccccg      973
tctctacaaa aaatacaaaa aaaaaaaaaa aaa      1006

<210> 101
<211> 1059
20 <212> DNA
    <213> Homo sapiens

    <220>
    <221> CDS
25 <222> 179..475

    <220>
    <221> sig_peptide
    <222> 179..295
30 <223> Von Heijne matrix
    score 4.14109371250204
    seq PSLIAGLFVGCLA/GY

    <400> 101
35 gtttttccag gagggagcgg cctttgctca gcgcgagacg gctgggccc gagtgggaca      60
   gcgctggtgc ggagactgct tccggactcc aggtaccgcg cttggcggca gctggcccca      120
   gacttctgtc ttttcagctg cagtgaaggc tcggggctgc agaattgcaa ccttgcca      178
   atg gac ctg atc ggt ttt ggt tat gca gcc ctc gtg aca ttt gga agc      226
Met Asp Leu Ile Gly Phe Gly Tyr Ala Ala Leu Val Thr Phe Gly Ser
40              -35              -30              -25
   att ttt gga tat aag cgg aga ggt ggt gtt ccg tct ttg att gct ggt      274
   Ile Phe Gly Tyr Lys Arg Arg Gly Gly Val Pro Ser Leu Ile Ala Gly
              -20              -15              -10
   ctt ttt gtt gga tgt ttg gcc gcc tat gga gct tac cgt gtc tcc aat      322
45 Leu Phe Val Gly Cys Leu Ala Gly Tyr Gly Ala Tyr Arg Val Ser Asn
              -5              1              5
   gac aaa cga gat gta aaa gtg tca ctg ttt aca gct ttc ttc ctg gct      370
   Asp Lys Arg Asp Val Lys Val Ser Leu Phe Thr Ala Phe Phe Leu Ala
10              15              20              25
50 acc ata atg ggt gtg aga ttt aag agg tcc aag aaa ata atg cct gct      418
   Thr Ile Met Gly Val Arg Phe Lys Arg Ser Lys Lys Ile Met Pro Ala
              30              35              40
   ggt ttg gtt gca ggt tta agc ctc atg atg atc ctg aga ctt gtc ttg      466
   Gly Leu Val Ala Gly Leu Ser Leu Met Met Ile Leu Arg Leu Val Leu
55              45              50              55
   ttg ctg ctc tgagcatctg gaggaacaga aaactaagtt catgtcatcc      515
   Leu Leu Leu
60
   tgctgtaatg ggcagagcat attttttttg tattttaaag ataaacttca atatggaatg      575
60 ctagaaacac aaatagcact gtcacctcta atatgaacat tagtttgagg tagttttttt      635
   ctaaagcaaa aatttttaact gtttttcta tgtcaagcac tattttcatt aaaagtgtct      695
   aatgaatcat gatatactct tccatttggt gtgtctatatt tttatatatt tgggtatttt      755
   tgaaaattcc aaatactcat gtctcaagta agcttaaact acaacttgtc acataaagga      815

```

```

      agtcttaagt ggagttcaca gaatgataat gtatctatatt gtcatttgtg ttatatattga      875
      aattattaga aattatgctt tttccatttt aattgtattg ctgccagtgc tattttttttc      935
      tttaaaaaat tttattctta gcacactggt atgtcctaac tgaatgtatt cagtattcaa      995
      ataaaagaca ttttgggtcca aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa     1055
5      aaaa                                     1059

      <210> 102
      <211> 514
      <212> DNA
10      <213> Homo sapiens

      <220>
      <221> CDS
      <222> 34..327
15

      <220>
      <221> sig_peptide
      <222> 34..162
      <223> Von Heijne matrix
20      score 5.69273078757386
      seq LGDALLFLRPAGS/CA

      <400> 102
25      agagacctat gggggttcgcc tgaagccccc gga atg tgt gag aca ctt ctt act      54
                                     Met Cys Glu Thr Leu Leu Thr
                                     -40
      agt aaa tgg gct tca gta tcc ccc atc cct gca ctc ctg cag gaa ggt      102
      Ser Lys Trp Ala Ser Val Ser Pro Ile Pro Ala Leu Leu Gln Glu Gly
      -35                                -30                                -25
30      gag aat cgg gac agt cgc agg ctg gga gac gct ctg ctt ttc ctg cgt      150
      Glu Asn Arg Asp Ser Arg Arg Leu Gly Asp Ala Leu Leu Phe Leu Arg
      -20                                -15                                -10                                -5
      cct gct ggg agc tgc gcg ctc cag gta tcc tgg cct gcc gcc cta gcc      198
      Pro Ala Gly Ser Cys Ala Leu Gln Val Ser Trp Pro Ala Ala Leu Ala
35      1                                5                                10
      ggc cca agg agc cac aca gga cag ttg acc caa cac ttc tgc cac ctg      246
      Gly Pro Arg Ser His Thr Gly Gln Leu Thr Gln His Phe Cys His Leu
      15                                20                                25
40      aag aac gac acc tgc att cct cca tct ctg gga cca cca agg aac tca      294
      Lys Asn Asp Thr Cys Ile Pro Pro Ser Leu Gly Pro Pro Arg Asn Ser
      30                                35                                40
      ggg agc ttg gaa tct ctc aga tca aaa aga tac tgactcatcg gatagccatg      347
      Gly Ser Leu Glu Ser Leu Arg Ser Lys Arg Tyr
      45                                50                                55
45      gcactctgaa aacggccttc cttgtgtgta cattatttgc aacaagcaac aagtttataa      407
      gcactttggt aaaattgcat gtgaggggta aaatattaaa gtcagtgcgt caacttgaaa      467
      taaatgatga gttattgatt actgctaaag aaaaaaaaaa aaaaaaa      514

      <210> 103
50      <211> 1158
      <212> DNA
      <213> Homo sapiens

      <220>
55      <221> CDS
      <222> 303..953

      <220>
      <221> sig_peptide
60      <222> 303..359
      <223> Von Heijne matrix
      score 5.47911600153114
      seq LCCSGCVPSLCCS/SY

```

	<400>	103	
	aaaaacttcc gccgccgcgt ccgcccgcctc cggaactaaa cgggggtgagg tcacattcgg	60	
	ttatctctaa cgtttgaaaa cgatggagct aacaccatt atggagatta accacttttc	120	
5	atcagggtttt taacttaagt cgtgaggaat acaacggtga acacaagatt cattttat	180	
	tcatacccat gggacgtatc ctgttggtga gttctctggg tcagacctct gaagacttct	240	
	cagatggatc ctagtctctg ggcttgccct gaaattactc gctgctcagg gagagagttg	300	
	aa atg gtt ggc atc ctc cca ctc tgt tgc tcc ggc tgt gtc ccc tcg	347	
	Met Val Gly Ile Leu Pro Leu Cys Cys Ser Gly Cys Val Pro Ser		
10	-15 -10 -5		
	ctc tgt tgt tcc agc tat gtc ccc tct gtt gct cca act gca gct cat	395	
	Leu Cys Cys Ser Ser Tyr Val Pro Ser Val Ala Pro Thr Ala Ala His		
	1 5 10		
	tct gtt aga gtt cct cat tca gct ggt cac tgt ggc cag agg gtg ttg	443	
15	Ser Val Arg Val Pro His Ser Ala Gly His Cys Gly Gln Arg Val Leu		
	15 20 25		
	gcc tgc tcc ctt cct caa gta ttc tta aag cca tgg att ttt gtg gag	491	
	Ala Cys Ser Leu Pro Gln Val Phe Leu Lys Pro Trp Ile Phe Val Glu		
	30 35 40		
20	cat ttt tct tcc tgg ctc tcc ctt gag tta ttt tcc ttt ctt cgc tat	539	
	His Phe Ser Ser Trp Leu Ser Leu Glu Leu Phe Ser Phe Leu Arg Tyr		
	45 50 55 60		
	ctt ggg act ctt ctt tgt gct tgc gga cat cgg ttg aga gaa gga cga	587	
	Leu Gly Thr Leu Leu Cys Ala Cys Gly His Arg Leu Arg Glu Gly Arg		
25	65 70 75		
	ctt ctt cct tgt ctc ctt ggt gtt ggc tgc tgg ttg ctc ttc aac aac	635	
	Leu Leu Pro Cys Leu Leu Gly Val Gly Ser Trp Leu Leu Phe Asn Asn		
	80 85 90		
	tgg act gga ggc tct tgg ttt tct ctt cat ctt caa caa gtc agt ctc	683	
30	Trp Thr Gly Ser Trp Phe Ser Leu His Leu Gln Gln Val Ser Leu		
	95 100 105		
	tct caa ggg tct cac gtt gca gca ttc tta cca gag gcc att ggg cct	731	
	Ser Gln Gly Ser His Val Ala Ala Phe Leu Pro Glu Ala Ile Gly Pro		
	110 115 120		
35	gga gtt cca gtt cca gtg tct gga gag tcc acc tca gct cag caa tct	779	
	Gly Val Pro Val Pro Val Ser Gly Glu Ser Thr Ser Ala Gln Gln Ser		
	125 130 135 140		
	cat gcc ggt tgg caa ttg tca gca gaa gcc gat gcc tgc cca tca gtt	827	
	His Ala Gly Trp Gln Leu Ser Ala Glu Ala Asp Ala Cys Pro Ser Val		
40	145 150 155		
	ctt tac tct gag gtg tta gag tgg aat aaa aat ata aat act tat act	875	
	Leu Tyr Ser Glu Val Leu Glu Trp Asn Lys Asn Ile Asn Thr Tyr Thr		
	160 165 170		
	agt ttt cat gac ttc tgc tta ata ttg ggt att ttt ktt gtt ttg ttt	923	
45	Ser Phe His Asp Phe Cys Leu Ile Leu Gly Ile Phe Xaa Val Leu Phe		
	175 180 185		
	tgt ttt ggc ggt gat agg ctt acc tta cat taaaccaggc ctttagccttt	973	
	Cys Phe Gly Gly Asp Arg Leu Thr Leu His		
	190 195		
50	ctgtggcttt gttatggcaa agcctcatat tactctctag tctgggttcag caggacagtc	1033	
	aggtccacac ctggggctgt ttgttttcta cgtttacctc aacataagggt accttatcat	1093	
	tgtcagcctt catctcctga tccaaaataa aataaaatgc cacagggtcaa aaaaaaaaaa	1153	
	aaaaa	1158	
55	<210> 104		
	<211> 1563		
	<212> DNA		
	<213> Homo sapiens		
60	<220>		
	<221> CDS		
	<222> 97..645		

```

<220>
<221> sig_peptide
<222> 97..156
<223> Von Heijne matrix
5   score 8.42885652997473
    seq AVVGCLLVPPAEA/NK

<220>
<221> misc_feature
10  <222> 972
    <223> n=a, g, c or t

<400> 104
15  aatagaagct aggagagggc ggggacaact gggctcttttg cggtctgcagc gggctttagtag      60
    gtgtccggct ttgctggccc agcaagcctg ataagc atg aag ctc tta tct ttg      114
                                   Met Lys Leu Leu Ser Leu
                                   -20                               -15

    gtg gct gtg gtc ggg tgt ttg ctg gtg ccc cca gct gaa gcc aac aag      162
    Val Ala Val Val Gly Cys Leu Leu Val Pro Pro Ala Glu Ala Asn Lys
    20  -10                               -5                               1

    agt tct gaa gat atc cgg tgc aaa tgc atc tgt cca cct tat aga aac      210
    Ser Ser Glu Asp Ile Arg Cys Lys Cys Ile Cys Pro Pro Tyr Arg Asn
                                   5                               10                               15

    atc agt ggg cac att tac aac cag aat gta tcc cag aag gac tgc aac      258
    Ile Ser Gly His Ile Tyr Asn Gln Asn Val Ser Gln Lys Asp Cys Asn
    25  20                               25                               30

    tgc ctg cac gtg gtg gag ccc atg cca gtg cct ggc cat gac gtg gag      306
    Cys Leu His Val Val Glu Pro Met Pro Val Pro Gly His Asp Val Glu
    35  35                               40                               45                               50

    gcc tac tgc ctg ctg tgc gag tgc agg tac gag gag cgc agc acc acc      354
    Ala Tyr Cys Leu Leu Cys Glu Cys Arg Tyr Glu Glu Arg Ser Thr Thr
    30  55                               60                               65

    acc atc aag gtc atc att gtc atc tac ctg tcc gtg gtg ggt gcc ctg      402
    Thr Ile Lys Val Ile Ile Val Ile Tyr Leu Ser Val Val Gly Ala Leu
    35  70                               75                               80

    ttg ctc tac atg gcc ttc ctg atg ctg gtg gac cct ctg atc cga aag      450
    Leu Leu Tyr Met Ala Phe Leu Met Leu Val Asp Pro Leu Ile Arg Lys
    85                               90                               95

    ccg gat gca tac act gag caa ctg cac aat gag gag gag aat gag gat      498
    Pro Asp Ala Tyr Thr Glu Gln Leu His Asn Glu Glu Glu Asn Glu Asp
    40  100                              105                              110

    gct cgc tct atg gca gca gct gct gca tcc ctc ggg gga ccc cga gca      546
    Ala Arg Ser Met Ala Ala Ala Ala Ala Ser Leu Gly Gly Pro Arg Ala
    45  115                              120                              125                              130

    aac aca gtc ctg gag cgt gtg gaa ggt gcc cag cag cgg tgg aag ctg      594
    Asn Thr Val Leu Glu Arg Val Glu Gly Ala Gln Gln Arg Trp Lys Leu
    135                              140                              145

    cag gtg cag gag cag cgg aag aca gtc ttc gat cgg cac aag atg ctc      642
    Gln Val Gln Glu Gln Arg Lys Thr Val Phe Asp Arg His Lys Met Leu
    50  150                              155                              160

    agc tagatgggct ggtgtggttg ggtcaaggcc ccaacaccat ggctgccagc      695
    Ser

    ttccaggctg gacaaagcag ggggctactt ctcccttccc tcggttccag tcttcccttt      755
    aaaagcctgt ggcatttttc ctcccttctcc ctaacttttag aaatgttgta cttggctatt      815
    55  ttgattaggg aagaggggatg tggctctctga tctctgttgt cttcttgggt ctttgggggt      875
    gaagggaggg ggaagggcagg ccacaaaggg aatggagaca ttcgaggcgg cctcaggagt      935
    ggatgcgacg ttgtctctcc tkggcctccc actcttngcc gccttcagc tctgagtctt      995
    gggaatgttg ttacccttgg aagataaagy ctgggtcttc aggaactcag tgtctggggag      1055
    gaaagcatgg ccagcatttc agcatgtgtt cctttctgca gtggttctta tcaccacctc      1115
    60  cctcccagcc ccagcgccctc agccccagcc ccagctccag ccctgaggac agctctgatg      1175
    ggagagctgg gccccctgag cccactgggt cttcagggtg cactggaagc tgggtgttcgc      1235
    tgtcccctgt gcacttctcg cactggggca tggagtgcc atgcatactc tgctgccggt      1295
    ccctcacct gcacttgagg ggtctgggca gtccctcctc tccccagtgt ccacagtcac      1355

```

```

tgagccagac ggtcgggttg aacatgagac tcgaggctga gcgtggatct gaacaccaca 1415
gcccctgtac ttgggttgcc tcttgtccct gaacttcgtt gtaccagtgc atggagagaa 1475
aattttgtcc tcttgtctta gagttgtgtg taaatcaagg aagccatcat taaattgttt 1535
tatttctctc taaaaaaaaa aaaaaaaaaa 1563
5
<210> 105
<211> 1621
<212> DNA
<213> Homo sapiens
10
<220>
<221> CDS
<222> 80..820
15
<220>
<221> sig_peptide
<222> 80..118
<223> Von Heijne matrix
      score 5.76690322882439
20      seq MLVLRSA LTRALA/SR

<400> 105
acctttccac tcgggaaacc ttcagaggag tctcagaaag gacacggctg gctgcttttc 60
tcagcgccga agccgcgcc atg ctc gtc ctc aga agc gcc ctg act cgg gcg 112
25      Met Leu Val Leu Arg Ser Ala Leu Thr Arg Ala
      -10 -5
ctg gcc tca cgg acg ctg gcg cct cag atg tgc tca tct ttt gct acg 160
Leu Ala Ser Arg Thr Leu Ala Pro Gln Met Cys Ser Ser Phe Ala Thr
      1 5 10
30 gga ccc aga caa tac gat gga ata ttc tat gaa ttt cgt tct tat tac 208
Gly Pro Arg Gln Tyr Asp Gly Ile Phe Tyr Glu Phe Arg Ser Tyr Tyr
      15 20 25 30
ctt aag ccc tca aag atg aat gag ttc ctg gaa aat ttt gag aaa aac 256
Leu Lys Pro Ser Lys Met Asn Glu Phe Leu Glu Asn Phe Glu Lys Asn
      35 40 45
35 gct cat ctt cgg aca gct cac tct gaa ttg gtt gga tac tgg agt gta 304
Ala His Leu Arg Thr Ala His Ser Glu Leu Val Gly Tyr Trp Ser Val
      50 55 60
40 gaa ttt gga ggc aga atg aat aca gtg ttt cat att tgg aag tat gat 352
Glu Phe Gly Gly Arg Met Asn Thr Val Phe His Ile Trp Lys Tyr Asp
      65 70 75
aat ttt gct cat cga act gaa gtt cag aaa gcc ttg gcc aaa gat aag 400
Asn Phe Ala His Arg Thr Glu Val Gln Lys Ala Leu Ala Lys Asp Lys
      80 85 90
45 gaa tgg caa gaa caa ttc ctc att cca aat ttg gct ctc att gat aaa 448
Glu Trp Gln Glu Gln Phe Leu Ile Pro Asn Leu Ala Leu Ile Asp Lys
      95 100 105 110
caa gag agt gag att act tat ctg gta cca tgg tgc aaa tta gaa aaa 496
Gln Glu Ser Glu Ile Thr Tyr Leu Val Pro Trp Cys Lys Leu Glu Lys
      115 120 125
50 cct cca aaa gaa gga gtc tat gaa ctg gcc act ttt cag atg aaa cct 544
Pro Pro Lys Glu Gly Val Tyr Glu Leu Ala Thr Phe Gln Met Lys Pro
      130 135 140
55 ggt ggg cca gct ctg tgg ggt gat gca ttt aaa agg gca gtt cat gct 592
Gly Gly Pro Ala Leu Trp Gly Asp Ala Phe Lys Arg Ala Val His Ala
      145 150 155
cat gtc aat cta ggc tac aca aaa cta gtt gga gtg ttc cac aca gag 640
His Val Asn Leu Gly Tyr Thr Lys Leu Val Gly Val Phe His Thr Glu
      160 165 170
60 tac gga gca ctc aac aga gtt cat gtt ctt tgg tgg aat gag agt gca 688
Tyr Gly Ala Leu Asn Arg Val His Val Leu Trp Trp Asn Glu Ser Ala
      175 180 185 190
gat agt cgt gca gct ggg aga cat aag tcc cat gag gat ccc aga gtt 736

```

```

Asp Ser Arg Ala Ala Gly Arg His Lys Ser His Glu Asp Pro Arg Val
195 200 205
gtg gca gct gtt cgg gaa agt gtc aac tac cta gta tct cag cag aat 784
Val Ala Ala Val Arg Glu Ser Val Asn Tyr Leu Val Ser Gln Gln Asn
5 210 215 220
atg ctt ctg att cct aca tcg ttt tca cca ctg aaa tagttttcta 830
Met Leu Leu Ile Pro Thr Ser Phe Ser Pro Leu Lys
225 230
ctgaaataca aaacattttca ttaactgcta taggatctct ctgctaattgg tgcttaaatt 890
10 ctcccaagag gttctcactt ttatttgaag gaggtggtaa gtttaatttgc tatgtttctt 950
gcattatgaa ggctacatct gtgctttgta agtaccactt caaaaaatag ttctgtttac 1010
tttctgcatg gtatttcagt gtctgtcata cattaaaaat acttgtcact gttttaagat 1070
cttgactctt cttttgtttc agaatagctc ttctactgta ttctgacaac tctttgcttt 1130
atagcatttt gttgtattca aatgataatg gtagcatttc catgcttggtg acagcatttt 1190
15 taagttatta atatatttta tcaacctttc catcatgtct gttttcctgg ttttttttgg 1250
ttgttttttg accagtaaaa tttattttgt aataccaaat aggatttaag aaaattaacg 1310
tatttcttta ctatggaaaa ccacattgtc atttgtgaca tcatctatat taaatatggg 1370
tttcacatta gttatttgtc acttacttgg aaaatgatgc tgtaggtcc tggattataa 1430
aatctagaaa agacttggtg gtttatgtgc tgaaatgtct ttatttataa ttaattttaa 1490
20 ctactattta ctttatttcg gatcctgttt aacaaagata cttgagacat ccatttgttt 1550
taatgaaatc tgtatggata tggaaatgct tgcctaata aaagcctaca tatacaaaaa 1610
aaaaaaaaa a 1621

<210> 106
25 <211> 557
<212> DNA
<213> Homo sapiens

<220>
30 <221> CDS
<222> 77..388

<220>
<221> sig_peptide
35 <222> 77..217
<223> Von Heijne matrix
score 4.57105404339594
seq FLYLTLNQSCIFA/NY

40 <400> 106
aacaccctcc ctggaccctc tgcctggagg acggggaatc acagcagctg gtttgggggtg 60
cctcccaaac caaaag atg ttc tct ccg cgc caa gct ttg acg ccc gac ccc 112
Met Phe Ser Pro Arg Gln Ala Leu Thr Pro Asp Pro
-45 -40
45 ctg cac tct ccc gcc tac tca ccg gtc cta ggg ggt tgg tcc cgc ttt 160
Leu His Ser Pro Ala Tyr Ser Pro Val Leu Gly Trp Ser Arg Phe
-35 -30 -25 -20
cgt agt gtg gat ttt cgt ttc ctc tac ttg act cta aat caa tcc tgt 208
Arg Ser Val Asp Phe Arg Phe Leu Tyr Leu Thr Leu Asn Gln Ser Cys
-15 -10 -5
50 ata ttc gca aac tac aaa gag gcg cat gca aat aga tac tgt act gag 256
Ile Phe Ala Asn Tyr Lys Glu Ala His Ala Asn Arg Tyr Cys Thr Glu
1 5 10
ggc aga tac acg cgc gag atc cag agg ctt aca tcc cca gcc gct tgg 304
55 Gly Arg Tyr Thr Arg Glu Ile Gln Arg Leu Thr Ser Pro Ala Ala Trp
15 20 25
ccc acc aga gac aag aac agg atg ata agc aat gga atg gca ttg aac 352
Pro Thr Arg Asp Lys Asn Arg Met Ile Ser Asn Gly Met Ala Leu Asn
30 35 40 45
60 tct cct gct gaa gga ctt gca ttt caa tgt aga ttc tgaggctggg 398
Ser Pro Ala Glu Gly Leu Ala Phe Gln Cys Arg Phe
50 55
tgaaaacttc tctgtcacct ttactacagc attctcaccc atttatattt ctttcccctt 458

```

```

ctacatctct attactgttg cactatgtta tgcattacac catggcaaaa ttaatcaatt 518
aatacaataa aagcttaatt ttaaaaaaaaa aaaaaaaaaa 557

<210> 107
5 <211> 600
  <212> DNA
  <213> Homo sapiens

<220>
10 <221> CDS
   <222> 139..513

<220>
15 <221> sig_peptide
   <222> 139..201
   <223> Von Heijne matrix
       score 5.86857787719223
       seq IVMGVQVVGRAFA/RA

20 <400> 107
   gaggcaatgc gcatgcccag cgccgtatcg cgcacgctct ctgcgggcttt ccttgacctc 60
   tgaccgcccg accacgcttg atccccggcc gcggggccag gaagtcggag tttgagcccc 120
   ggaggcagag cggctgcc atg gcc aag tac ctg gcc cag atc att gtg atg 171
                               Met Ala Lys Tyr Leu Ala Gln Ile Ile Val Met
25                               -20                               -15
   ggc gtg cag gtg gtg ggc agg gcc ttt gca cgg gcc ttg cgg cag gag 219
   Gly Val Gln Val Val Gly Arg Ala Phe Ala Arg Ala Leu Arg Gln Glu
   -10                               -5                               1                               5
   ttt gca gcc agc cgg gcc gca gct gat gcc cga gga cgc gct gga cac 267
30 Phe Ala Ala Ser Arg Ala Ala Ala Asp Ala Arg Gly Arg Ala Gly His
                               10                               15                               20
   cgg tct gca gcc gct tcc aac ctc tcc ggc ctc agc ctc cag gag gca 315
   Arg Ser Ala Ala Ala Ser Asn Leu Ser Gly Leu Ser Leu Gln Glu Ala
                               25                               30                               35
35 cag cag att ctc aac gtg tcc aag ctg agc cct gag gag gtc cag aag 363
   Gln Gln Ile Leu Asn Val Ser Lys Leu Ser Pro Glu Glu Val Gln Lys
   40                               45                               50
   aac tat gaa cac tta ttt aag gtg aat gat aaa tcc gtg ggt ggc tcc 411
   Asn Tyr Glu His Leu Phe Lys Val Asn Asp Lys Ser Val Gly Gly Ser
40 55                               60                               65                               70
   ttc tac ctg cag tca aag gtg gtc cgc gca aag gag cgc ctg gat gag 459
   Phe Tyr Leu Gln Ser Lys Val Val Arg Ala Lys Glu Arg Leu Asp Glu
                               75                               80                               85
   gaa ctc aaa atc cag gcc cag gag gac aga gaa aaa ggg cag atg ccc 507
45 Glu Leu Lys Ile Gln Ala Gln Glu Asp Arg Glu Lys Gly Gln Met Pro
   90                               95                               100
   cat acg tgactgctcg gctccccccg cccaccccg cgcctctaatt ttatagcttg 563
   His Thr
   gtaataaatt tcttttctac aaaaaaaaaa aaaaaaaa 600

50 <210> 108
   <211> 1129
   <212> DNA
   <213> Homo sapiens

55 <220>
   <221> CDS
   <222> 81..986

60 <220>
   <221> sig_peptide
   <222> 81..134
   <223> Von Heijne matrix

```

score 5.03543461931947

seq ITLLGLAVNVVTT/LV

<400> 108

```

5  acagcgcggc gggcgtctcg ctgctcgagc cgccgctgca gctctactgg acctggctgc      60
   tccagtggat cccgctctgg atg gcc ccc aac tcc atc acc ctg ctg ggg ctc      113
                                   Met Ala Pro Asn Ser Ile Thr Leu Leu Gly Leu
                                   -15                               -10
10  gcc gtc aac gtg gtc acc acg ctc gtg ctc atc tcc tac tgt ccc acg      161
   Ala Val Asn Val Val Thr Thr Leu Val Leu Ile Ser Tyr Cys Pro Thr
                                   -5                               5
   gcc acc gaa gag gca cca tac tgg aca tac ctt tta tgt gca ctg gga      209
   Ala Thr Glu Glu Ala Pro Tyr Trp Thr Tyr Leu Leu Cys Ala Leu Gly
   10                               15                               20       25
15  ctt ttt att tac cag tca ctg gat gct att gat ggg aaa caa gcc aga      257
   Leu Phe Ile Tyr Gln Ser Leu Asp Ala Ile Asp Gly Lys Gln Ala Arg
                                   30                               35       40
   aga aca aac tct tgt tcc cct tta ggg gag ctc ttt gac cat ggc tgt      305
   Arg Thr Asn Ser Cys Ser Pro Leu Gly Glu Leu Phe Asp His Gly Cys
   20                               45                               50       55
   gac tct ctt tcc aca gta ttt atg gca gtg gga gct tca att gcc gct      353
   Asp Ser Leu Ser Thr Val Phe Met Ala Val Gly Ala Ser Ile Ala Ala
                                   60                               65       70
25  cgc tta gga act tat cct gac tgg ttt ttt ttc tgc tct ttt att ggg      401
   Arg Leu Gly Thr Tyr Pro Asp Trp Phe Phe Phe Cys Ser Phe Ile Gly
   75                               80                               85
   atg ttt gtg ttt tat tgc gct cat tgg cag act tat gtt tca ggc atg      449
   Met Phe Val Phe Tyr Cys Ala His Trp Gln Thr Tyr Val Ser Gly Met
   90                               95                               100      105
30  ttg aga ttt gga aaa gtg gat gta act gaa att cag ata gct tta gtg      497
   Leu Arg Phe Gly Lys Val Asp Val Thr Glu Ile Gln Ile Ala Leu Val
                                   110                              115      120
   att gtc ttt gtg ttg tct gca ttt gga gga gca aca atg tgg gac tat      545
   Ile Val Phe Val Leu Ser Ala Phe Gly Gly Ala Thr Met Trp Asp Tyr
   125                              130                              135
35  acg ggc acc agt gtc ttg tca cct gga ctc cac ata gga cta att att      593
   Thr Gly Thr Ser Val Leu Ser Pro Gly Leu His Ile Gly Leu Ile Ile
                                   140                              145      150
   ata ctg gca ata atg atc tat aaa aag tca gca act gat gtg ttt gaa      641
   Ile Leu Ala Ile Met Ile Tyr Lys Lys Ser Ala Thr Asp Val Phe Glu
   155                              160                              165
   aag cat cct tgt ctt tat atc cta atg ttt gga tgt gtc ttt gct aaa      689
   Lys His Pro Cys Leu Tyr Ile Leu Met Phe Gly Cys Val Phe Ala Lys
   170                              175                              180      185
45  gtc tca caa aaa tta gtg gta gct cac atg acc aaa agt gaa cta tat      737
   Val Ser Gln Lys Leu Val Val Ala His Met Thr Lys Ser Glu Leu Tyr
                                   190                              195      200
   ctt caa gac act gtc ttt ttg ggg cca ggt ctt ttg ttt tta gac cag      785
   Leu Gln Asp Thr Val Phe Leu Gly Pro Gly Leu Leu Phe Leu Asp Gln
   205                              210                              215
50  tac ttt aat aac ttt ata gac gaa tat gtt gtt cta tgg atg gca atg      833
   Tyr Phe Asn Asn Phe Ile Asp Glu Tyr Val Val Leu Trp Met Ala Met
                                   220                              225      230
   gtg att tct tca ttt gat atg gtg ata tac ttt agt gct ttg tgc ctg      881
   Val Ile Ser Ser Phe Asp Met Val Ile Tyr Phe Ser Ala Leu Cys Leu
   235                              240                              245
   caa att tca aga cac ctt cat cta aat ata ttc aag act gca tgt cat      929
   Gln Ile Ser Arg His Leu His Leu Asn Ile Phe Lys Thr Ala Cys His
   250                              255                              260      265
60  caa gca cct gaa cag gtt caa gtt ctt tct tca aag agt cat cag aat      977
   Gln Ala Pro Glu Gln Val Gln Val Leu Ser Ser Lys Ser His Gln Asn
                                   270                              275      280
   aac atg gat tgaagagact tccgaacact tgctatctct tgctgctgct      1026

```


Asn Met Asp
gtttcatgga aggagatatt aaacatttgt ttaattttta tttaagtgtt atacctattt 1086
cagcaaataa aatatttcac tgcttgaaaa aaaaaaaaaa aaa 1129

5 <210> 109
<211> 778
<212> DNA
<213> Homo sapiens

10 <220>
<221> CDS
<222> 266..586

<220>
15 <221> sig_peptide
<222> 266..307
<223> Von Heijne matrix
score 4.534746808071
seq ILVTVPGVCPAQC/CW

20
<400> 109
tagatgttag aattgggtatt tggtcttgct ttttggttgc gatggagtta tataactaagt 60
tacttatact aaggcattag tagtctcata tctgaggagc aattgtattt ttagttcagc 120
taaattaatg cctcttttta aataactaact tgtactactt ttgtggctgt gaatggatc 180
25 ttttattgaa ctgaggcagc ttttaaaaga cttgcctgat catttagagc actcccattg 240
agggttaaatt agacttgaat ctgta atg att ctc gta act gtt cct ggt gtg 292
Met Ile Leu Val Thr Val Pro Gly Val
-10
tgt cca gca caa tgt tgc tgg gca gag cag agg ggc aga ggc tca ggt 340
30 Cys Pro Ala Gln Cys Cys Trp Ala Glu Gln Arg Gly Arg Gly Ser Gly
-5 1 5 10
atg tac ttc att gac aag tgg gca agg cca tcc tgg gta cca cat tgg 388
Met Tyr Phe Ile Asp Lys Trp Ala Arg Pro Ser Trp Val Pro His Trp
15 20 25
35 ctt aat gat ctc ttc att gtg aag tcc ggc tac ctc gtt tgc ata aga 436
Leu Asn Asp Leu Phe Ile Val Lys Ser Gly Tyr Leu Val Cys Ile Arg
30 35 40
act aca gta atc agg caa ggc att gtc aga att ggg agg aat aaa atc 484
Thr Thr Val Ile Arg Gln Gly Ile Val Arg Ile Gly Arg Asn Lys Ile
40 45 50 55
agt gag tct gga agg agt gct ctg tat aca att gca aag aac aaa atg 532
Ser Glu Ser Gly Arg Ser Ala Leu Tyr Thr Ile Ala Lys Asn Lys Met
60 65 70 75
gtc atc ttt aag gta cct gat tgc atg cac tta aat gca gat tat ttt 580
45 Val Ile Phe Lys Val Pro Asp Cys Met His Leu Asn Ala Asp Tyr Phe
80 85 90
gga gtt tgaaaaggga ctattaatga aatctttctt ttccctcctt tctctttttc 636
Gly Val
ccttccccgc cactgattca gtgagctgga gattggatca cagccgaagg agtaaagggtg 696
50 ctgcaatgat gttagctgtg gccactgtgg atttttcgca agaacattaa taaactaaaa 756
atttcaaaaa aaaaaaaaaa aa 778

<210> 110
<211> 1301
55 <212> DNA
<213> Homo sapiens

<220>
<221> CDS
60 <222> 59..745

<220>
<221> sig_peptide

<222> 59..160

<223> Von Heijne matrix

score 5.94384548075359

seq LGAAALALLLANT/DV

5

<400> 110

	attcaaaacc	aggctgaaga	ttggaaggaa	gttggccagc	ctcggctgca	ggacagaa	58
	atg tct ttc ctc cag gac cca agt ttc ttc acc atg ggg atg tgg tcc	106					
	Met Ser Phe Leu Gln Asp Pro Ser Phe Phe Thr Met Gly Met Trp Ser						
10	-30	-25	-20				
	att ggt gca gga gcc ctg ggg gct gct gcc ttg gca ttg ctg ctt gcc	154					
	Ile Gly Ala Gly Ala Leu Gly Ala Ala Leu Ala Leu Leu Ala						
	-15	-10	-5				
	aac aca gac gtg ttt ctg tcc aag ccc cag aaa gcg gcc ctg gag tac	202					
15	Asn Thr Asp Val Phe Leu Ser Lys Pro Gln Lys Ala Ala Leu Glu Tyr						
	1	5	10				
	ctg gag gat ata gac ctg aaa aca ctg gag aag gaa cca agg act ttc	250					
	Leu Glu Asp Ile Asp Leu Lys Thr Leu Glu Lys Glu Pro Arg Thr Phe						
	15	20	25				
20	aaa gca aag gag cta tgg gaa aaa aat gga gct gtg att atg gcc gtg	298					
	Lys Ala Lys Glu Leu Trp Glu Lys Asn Gly Ala Val Ile Met Ala Val						
	35	40	45				
	cgg agg cca ggc tgt ttc ctc tgt cga gag gaa gct gcg gat ctg tcc	346					
	Arg Arg Pro Gly Cys Phe Leu Cys Arg Glu Glu Ala Ala Asp Leu Ser						
25	50	55	60				
	tcc ctg aaa agc atg ttg gac cag ctg ggc gtc ccc ctc tat gca gtg	394					
	Ser Leu Lys Ser Met Leu Asp Gln Leu Gly Val Pro Leu Tyr Ala Val						
	65	70	75				
	gta aag gag cac atc agg act gaa gtg aag gat ttc cag cct tat ttc	442					
30	Val Lys Glu His Ile Arg Thr Glu Val Lys Asp Phe Gln Pro Tyr Phe						
	80	85	90				
	aaa gga gaa atc ttc ctg gat gaa aag aaa aag ttc tat ggt cca caa	490					
	Lys Gly Glu Ile Phe Leu Asp Glu Lys Lys Lys Phe Tyr Gly Pro Gln						
	95	100	105				
35	agg cgg aag atg atg ttt atg gga ttt atc cgt ctg gga gtg tgg tac	538					
	Arg Arg Lys Met Met Phe Met Gly Phe Ile Arg Leu Gly Val Trp Tyr						
	115	120	125				
	aac ttc ttc cga gcc tgg aac gga ggc ttc tct gga aac ctg gaa gga	586					
	Asn Phe Phe Arg Ala Trp Asn Gly Gly Phe Ser Gly Asn Leu Glu Gly						
40	130	135	140				
	gaa ggc ttc atc ctt ggg gga gtt ttc gtg gtg gga tca gga aag cag	634					
	Glu Gly Phe Ile Leu Gly Gly Val Phe Val Val Gly Ser Gly Lys Gln						
	145	150	155				
	ggc att ctt ctt gag cac cga gaa aaa gaa ttt gga gac aaa gta aac	682					
45	Gly Ile Leu Leu Glu His Arg Glu Lys Glu Phe Gly Asp Lys Val Asn						
	160	165	170				
	cta ctt tct gtt ctg gaa gct gct aag atg atc aaa cca cag act ttg	730					
	Leu Leu Ser Val Leu Glu Ala Ala Lys Met Ile Lys Pro Gln Thr Leu						
	175	180	185				
50	gcc tca gag aaa aaa tgattgtgtg aaactgccca gctcagggat aaccagggac	785					
	Ala Ser Glu Lys Lys						
	195						
	attcacctgt gttcatgga tgtattgttt ccactcgtgt ccctaaggag tgagaaaccc	845					
	atttatactc tactctcagt atggattatt aatgtatttt aatattctgt ttaggccac	905					
55	taaggcaaaa tagcccaaaa acaagactga caaaaatctg aaaaactaat gaggattatt	965					
	aagctaaaaac ctgggaaata ggagggttaa aattgactgc caggctgggt gcagtggctc	1025					
	acacctgtaa tcccagcact ttgggaggcc aagggtgagca agtcacttga ggtcgggagt	1085					
	tcgagaccag cctgagcaac atggcgaaac cccgtctcta ctaaaaatac aaaaatcacc	1145					
	cgggtgtggt ggcaggcacc tgtagtccca gctacccggg aggctgaggc aggagaatca	1205					
60	cttgaacctg ggaggtggag gttgcggtga gctgagatca caccactgta ttccagcctg	1265					
	ggtgactgag actctaacta aaaaaaaaaa aaaaaa	1301					

<210> 111

<211> 1300
 <212> DNA
 <213> Homo sapiens

5 <220>
 <221> CDS
 <222> 59..676

<220>
 10 <221> sig_peptide
 <222> 59..160
 <223> Von Heijne matrix
 score 5.94384548075359
 seq LGAAALALLLANT/DV

15
 <400> 111
 attcaaaacc aggctgaaga ttggaaggaa gttggccagc ctcggctgca ggacagaa 58
 atg tct ttc ctc cag gac cca agt ttc ttc acc atg ggg atg tgg tcc 106
 Met Ser Phe Leu Gln Asp Pro Ser Phe Phe Thr Met Gly Met Trp Ser
 20 att ggt gca gga gcc ctg ggg gct gct gcc ttg gca ttg ctg ctt gcc 154
 Ile Gly Ala Gly Ala Leu Gly Ala Ala Leu Ala Leu Leu Leu Ala
 -30 -25 -20
 aac aca gac gtg ttt ctg tcc aag ccc cag aaa gcg gcc ctg gag tac 202
 25 Asn Thr Asp Val Phe Leu Ser Lys Pro Gln Lys Ala Ala Leu Glu Tyr
 1 5 10
 ctg gag gat ata gac ctg aaa aca ctg gag aag gaa cca agg act ttc 250
 Leu Glu Asp Ile Asp Leu Lys Thr Leu Glu Lys Glu Pro Arg Thr Phe
 15 20 25 30
 30 aaa gca aag gag cta tgg gaa aaa aat gga gct gtg att atg gcc gtg 298
 Lys Ala Lys Glu Leu Trp Glu Lys Asn Gly Ala Val Ile Met Ala Val
 35 35 40 45
 cgg agg cca ggc tgt ttc ctc tgt cga gag gaa gct gcg gat ctg tcc 346
 Arg Arg Pro Gly Cys Phe Leu Cys Arg Glu Glu Ala Ala Asp Leu Ser
 50 55 60
 35 tcc ctg aaa agc atg ttg gac cag ctg ggc gtc ccc ctc tat gca gtg 394
 Ser Leu Lys Ser Met Leu Asp Gln Leu Gly Val Pro Leu Tyr Ala Val
 65 70 75
 gta aag gag cac atc agg act gaa gtg aag gat ttc cag cct tat ttc 442
 40 Val Lys Glu His Ile Arg Thr Glu Val Lys Asp Phe Gln Pro Tyr Phe
 80 85 90
 aaa gga gaa atc ttc ctg gat gaa aag aaa aag ttc tat ggt cca caa 490
 Lys Gly Glu Ile Phe Leu Asp Glu Lys Lys Lys Phe Tyr Gly Pro Gln
 95 100 105 110
 45 agg cgg aag atg atg ttt atg gga ttt atc cgt ctg gga gtg tgg tac 538
 Arg Arg Lys Met Phe Met Gly Phe Ile Arg Leu Gly Val Trp Tyr
 115 120 125
 aac ttc ttc cga gcc tgg aac gga ggc ttc tct gga aac ctg gaa gga 586
 Asn Phe Phe Arg Ala Trp Asn Gly Gly Phe Ser Gly Asn Leu Glu Gly
 130 135 140
 50 gaa ggc ttc atc ctt ggg gga gtt ttc gtg gtg gga tca gga agc agg 634
 Glu Gly Phe Ile Leu Gly Gly Val Phe Val Val Gly Ser Gly Ser Arg
 145 150 155
 gca ttc ttc ttg agc acc gag aaa aag aat ttg gag aca aag 676
 55 Ala Phe Phe Leu Ser Thr Glu Lys Lys Asn Leu Glu Thr Lys
 160 165 170
 taaacctact ttctgttctg gaagctgcta agatgatcaa accacagact ttggcctcag 736
 agaaaaaatg attgtgtgaa actgcccagc tcagggataa ccagggacat tcacctgtgt 796
 tcatgggatg tattgtttcc actcgtgtcc ctaaggagtg agaaacccat ttatactcta 856
 60 ctctcagtat ggattattaa tgtatttttaa tattctgttt aggccacta aggcaaaata 916
 gccccaaaac aagactgaca aaaatctgaa aaactaatga ggattattaa gctaaaacct 976
 gggaaatagg aggttttaaaa ttgactgcca ggctgggtgc agtggctcac acctgtaatc 1036
 ccagcacttt gggaggccaa ggtgagcaag tcacttgagg tcgggagttc gagaccagcc 1096

	tgagcaacat	ggcgaaaccc	cgtctctact	aaaaatacaa	aaatcacccg	ggtgtggtgg	1156										
	caggcacctg	tagtcccagc	taccggggag	gctgaggcag	gagaatcact	tgaacctggg	1216										
	aggtggaggt	tgcggtgagc	tgagatcaca	ccactgtatt	ccagcctggg	tgactgagac	1276										
	tctaactaaa	aaaaaaaaaa	aaaa				1300										
5																	
	<210> 112																
	<211> 1617																
	<212> DNA																
	<213> Homo sapiens																
10																	
	<220>																
	<221> CDS																
	<222> 15..278																
15																	
	<220>																
	<221> sig_peptide																
	<222> 15..146																
	<223> Von Heijne matrix																
	score 12.2610572403264																
20	seq PLFLLLLLGSVTA/DI																
	<400> 112																
	gagaggagag	gaga	atg	gcg	gcg	gaa	ggc	tgg	att	tgg	cgt	tgg	ggc	tgg	50		
			Met	Ala	Ala	Glu	Gly	Trp	Ile	Trp	Arg	Trp	Gly	Trp			
25							-40					-35					
	ggc	cgg	cgg	tgc	ctg	gga	agg	cct	ggg	ctt	ctc	ggc	ccc	ggc	cct	ggc	98
	Gly	Arg	Arg	Cys	Leu	Gly	Arg	Pro	Gly	Leu	Leu	Gly	Pro	Gly	Pro	Gly	
			-30				-25					-20					
	ccc	act	aca	cct	ctc	ttt	ctt	ctt	ttg	ttg	ttg	ggg	tct	gtg	act	gcg	146
30	Pro	Thr	Thr	Pro	Leu	Phe	Leu	Leu	Leu	Leu	Leu	Gly	Ser	Val	Thr	Ala	
		-15				-10					-5						
	gat	ata	act	gac	ggc	aac	att	gaa	cat	ctc	aag	cgg	gag	cat	tcg	ctc	194
	Asp	Ile	Thr	Asp	Gly	Asn	Ile	Glu	His	Leu	Lys	Arg	Glu	His	Ser	Leu	
	1				5					10				15			
35	att	aag	ccc	tac	caa	ggg	gtc	ggt	tcc	agc	tcc	ccc	tct	ggg	act	tcc	242
	Ile	Lys	Pro	Tyr	Gln	Gly	Val	Gly	Ser	Ser	Ser	Pro	Ser	Gly	Thr	Ser	
			20					25				30					
	agg	gca	gca	cta	tgc	tca	cga	gcc	agt	acg	tac	gtc	tgacc	ccctga			288
	Arg	Ala	Ala	Leu	Cys	Ser	Arg	Ala	Ser	Thr	Tyr	Val					
40		35					40										
	cgagcgcagc	aaagagggct	ctatctggaa	ccaccagccg	tgcttctctca	aagactggga										348	
	aatgcacgtc	cacttcaaag	tccacggcac	agggagaaga	aacctccatg	gagacggcat										408	
	cgccttgttg	tacaccgggg	accgcctcgt	gccagggcct	gtgttttgaa	gcaaagataa										468	
	cttccacggc	ttagccatct	tcctggacac	ctaccccaat	gatgagacca	ctgagcgcgt										528	
45	gttcccgtac	atctcggtga	tggtgaacaa	tggtcctctg	tcctacgacc	acagcaagga										588	
	tgggcgctgg	accgagctgg	cgggctgcac	ggctgacttc	cgcaaccgcg	atcacgacac										648	
	cttcctggct	gtgcgtact	cccggggccg	tctgacgggt	atgaccgacc	tggaggacaa										708	
	gaacgagtgg	aagaactgca	ttgacatcac	gggagtgcgc	ctgcccaccg	gctactactt										768	
	cggggcctcc	gccggcaccg	gcgacctgtc	tgacaatcat	gacatcatct	ccatgaagct										828	
50	gttccagctg	atggtggagc	acacgcccga	cgaggagagc	atcgactgga	ccaagatcga										888	
	gcccagcgtc	aacttcctca	agtcgcccga	agacaacgtg	gacgacccca	cggggaactt										948	
	ccgcagcggg	cccctgacgg	ggtggcgggt	gttcctgctg	ctgctgtgcg	ctctcctggg										1008	
	catcgttgct	tgcgccgtgg	tgggggccgt	ggtgttccag	aagcggcagg	agcggaaacaa										1068	
	gcgcttctac	tgagtggcgc	ctccggcggg	gcctgtccct	gggcccagga	gccaatgtga										1128	
55	actttttttt	taccgggatt	ataaaagaac	aacaagatga	ccttatttct	taactgtttc										1188	
	aaataaatga	ttaaagtatt	ttcatacatt	ttgcttcttg	cccagcaggg	acaggtggca										1248	
	gagccgaggc	ttaggtgtct	gcacccccca	cagctggaga	cggaggctct	cctggggctg										1308	
	gtgtctcagg	agcaggggtc	tgtgtctaca	gatgggctgt	ggcccctgca	ggcagctgtt										1368	
	gaacactgga	gggtcccccg	gaccacactg	gggtgggctc	ctgaggacgt	ggggaagtga										1428	
60	ttttgttttg	tggtgtgtgg	cacgtgtggc	gacggataag	gcctgaactg	gaaacccagg										1488	
	ccttcctgtt	caccctgagc	tgcttctctga	gacagatgct	caagtgaggc	tgcaggcgcg										1548	
	gtgtgggtgg	gccgagtgtg	accgtttgct	aaataaagtg	aaatacccaa	caaaaaaaaaa										1608	
	aaaaaaaaaa															1617	

<210> 113
 <211> 1634
 <212> DNA
 5 <213> Homo sapiens

 <220>
 <221> CDS
 <222> 167..619
 10
 <220>
 <221> sig_peptide
 <222> 167..262
 <223> Von Heijne matrix
 15 score 6.8501239662158
 seq LLSSCGLPPSTAS/AV

 <400> 113
 20 gtactttccga gagagattaa agattcaatg gaactctgcg tctctcatct ggaacccagg 60
 acacagaaca agggagggaa gaaaagctca gccttaaaca tagcaagggtg aaacctttgt 120
 cctgggggaat agtctggccc gctccttgga accacactca gactca atg gac tct 175
 Met Asp Ser
 -30
 25 gcc tca aat ccc acc aac ctt gtc agc acc tcc caa agg cac cgg ccc 223
 Ala Ser Asn Pro Thr Asn Leu Val Ser Thr Ser Gln Arg His Arg Pro
 -25 -20 -15
 ttg ctt tca tcc tgt ggc ctc cca cca agc act gcc tca gct gtg cgc 271
 Leu Leu Ser Ser Cys Gly Leu Pro Pro Ser Thr Ala Ser Ala Val Arg
 -10 -5 1
 30 agg cta tgc tcc agg gga gtg tta aaa gga tca aat gaa aga agg gat 319
 Arg Leu Cys Ser Arg Gly Val Leu Lys Gly Ser Asn Glu Arg Arg Asp
 5 10 15
 atg gaa tca ttt tgg aaa cta aat cgt tcc cca ggg tcg gac cga tac 367
 Met Glu Ser Phe Trp Lys Leu Asn Arg Ser Pro Gly Ser Asp Arg Tyr
 20 25 30 35
 35 ctg gag agc cgc gat gcc tct cga ctg agt ggc cgg gac ccc tcc tca 415
 Leu Glu Ser Arg Ala Ser Arg Leu Ser Gly Arg Asp Pro Ser Ser
 40 45 50
 tgg aca gtc gag gat gtg atg cag ttt gtc cgg gaa gct gat cct cag 463
 Trp Thr Val Glu Asp Val Met Gln Phe Val Arg Glu Ala Asp Pro Gln
 55 60 65
 ctt gga ccc cac gct gac ctg ttt cgc aaa cac gag atc gat ggc aag 511
 Leu Gly Pro His Ala Asp Leu Phe Arg Lys His Glu Ile Asp Gly Lys
 70 75 80
 45 gcc ctg ctg ctg ctg cgc agt gac atg atg atg aag tac atg ggc ctg 559
 Ala Leu Leu Leu Leu Arg Ser Asp Met Met Met Lys Tyr Met Gly Leu
 85 90 95
 aag ctg ggg cct gca ctc aag ctc tcc tac cac att gac cgg ctg aag 607
 Lys Leu Gly Pro Ala Leu Lys Leu Ser Tyr His Ile Asp Arg Leu Lys
 100 105 110 115
 50 cag ggc aag ttc tgaaccagga gaggcagcct agacaaccaa gtggcagcag 659
 Gln Gly Lys Phe
 gtgggggcat tcttctagga atgaggggca tcagcccacc ccaggcacct cagtgggggtt 719
 ccggggccacc tcaggactcc aagaggctgt gtggagccac cactcctagc cacagctgcc 779
 55 atgataagtc cttccatgaa ggactgagga gggagagtgg ggggtccaggg ctggtgctgc 839
 tcttcctca gctctgccgg ggctctaagg tccctctatt tatttctcaa ccttggtcgg 899
 cctctcacca ggagtttagg ctgaatgcct tccacgtgat ggaggaaaag gccaaactctg 959
 tctgtgtctt gctgtggcac cccatcgccc cacagctcgt accttctcac cagattcccc 1019
 tgaatccaaa ctsgtggtgc aaacctctac cttttttaca aaaagatctt attgttaatt 1079
 60 tattgtttct ggcacttggg caaacctctgt agttaatact cctccccmac actagacact 1139
 gggtttcagg aggagggaga ctgccctgct ttgggtcccca gagaggccct ctgcagatag 1199
 gcgtggcccc tcttcagagg acactaccct agggcacttt ctctttgagg tggagagacc 1259
 cataaagcct tgaccacatc actccatatg gggaggagaa ggatccctgt caccttctcc 1319

```

tctcttcaag gggccctttt gcagccctag gcctcatctg tgggaaggga gtccctggct 1379
tatactgccc ccaccacagc tccttgccct ggccagaact gctgtcgaag aaaatcaggc 1439
cggaaggcca agaaggcgct aagggggatg ggagggcagg ttttccaggc tggagtcggg 1499
tccaccact cgcctgtcca caggcttcct tgtaagcaag tcagcagcac agctactcac 1559
5 gctgccatct ggacttattt tatgtcaatc tgtttataaa taaaaaccaa tataggtaaa 1619
aaaaa aaaaa 1634

<210> 114
<211> 693
10 <212> DNA
    <213> Homo sapiens

    <220>
    <221> CDS
15 <222> 223..417

    <220>
    <221> sig_peptide
    <222> 223..270
20 <223> Von Heijne matrix
    score 4.19788230215007
    seq LACVRESTSVAWA/CK

    <400> 114
25 ttagggggcc tgtcaccag cacgtgcac gggggctgtc ccgggggtca ggggagggag 60
   gccagcgggc mgtgtcgggg tccgccccga ccccatccac gaccccgact cctatccgat 120
   cctatccccg gccccgctcg ggcctttccc cttgcgcctt ggctcggctg gctcgcagag 180
   cagtaagtgc gtacccgccc tccgaagccg ggcgtgcacg gg atg gca gag ttg 234
                                     Met Ala Glu Leu
30                                     -15
   gcg tgc gtg cgt gag tcc acc agt gtg gca tgg gca tgt aag gtg cgc 282
   Ala Cys Val Arg Glu Ser Thr Ser Val Ala Trp Ala Cys Lys Val Arg
                                     -10 -5 1
35 gga ggg act gca cct tct cca tca ggt gca gaa ggc cac gtc atg ctg 330
   Gly Gly Thr Ala Pro Ser Pro Ser Gly Ala Glu Gly His Val Met Leu
   5 10 15 20
   aac aag agc cga gaa gta gaa tcg cca gtg tca agc cgt cca cgt tgt 378
   Asn Lys Ser Arg Glu Val Glu Ser Pro Val Ser Ser Arg Pro Arg Cys
                                     25 30 35
40 ggg atg ccc act gtt ccc cca gga tca ctc aag acc ctg tgacttgagg 427
   Gly Met Pro Thr Val Pro Pro Gly Ser Leu Lys Thr Leu
                                     40 45
   tcactgatga gtggaccaag tgaagtccac aagatggctg ctgtggctcc aggcacacag 487
   tccacatgca aatccatcca gaggcaggaa ctgggaatag gcttggagggt ggccaggaca 547
45 gcaagtgggc tgtctgtata aacctccctt ccacttggga aggaaaatca ccccccaagt 607
   cgattttctg tccatcttat tgatcagaga gcgttataaa ttcacccatt aaataatctg 667
   gacaagggga aaaaaaaaaa aaaaaa 693

    <210> 115
50 <211> 784
    <212> DNA
    <213> Homo sapiens

    <220>
    <221> CDS
55 <222> 166..732

    <220>
    <221> sig_peptide
60 <222> 166..237
    <223> Von Heijne matrix
    score 6.60662787180923
    seq KMVHLLVLSGAWG/MQ

```

<400> 115
 attattggtt gggggaaacc cacgagggga cgcggccgag gagggtcgct gtccaccg 60
 gggcgtggga gtgaggtacc agattcagcc catttgccc cgacgcctct gttctcggaa 120
 5 tccgggtgct gcggattgag gtccccggttc ctaacggact gcaag atg gag gaa ggc 177
 Met Glu Glu Gly
 ggg aac cta gga ggc ctg att aag atg gtc cat cta ctg gtc ttg tca 225
 Gly Asn Leu Gly Gly Leu Ile Lys Met Val His Leu Leu Val Leu Ser
 -20 -15 -10 -5
 10 ggt gcc tgg ggc atg caa atg tgg gtg acc ttc gtc tca ggc ttc ctg 273
 Gly Ala Trp Gly Met Gln Met Trp Val Thr Phe Val Ser Gly Phe Leu
 1 5 10
 ctt ttc cga agc ctt ccc cga cat acc ttc gga cta gtg cag agc aaa 321
 Leu Phe Arg Ser Leu Pro Arg His Thr Phe Gly Leu Val Gln Ser Lys
 15 15 20 25
 ctc ttc ccc ttc tac ttc cac atc tcc atg ggc tgt gsc ttc atc aac 369
 Leu Phe Pro Phe Tyr Phe His Ile Ser Met Gly Cys Xaa Phe Ile Asn
 30 35 40
 20 ctc tgc atc ttg gct tca cag cat gct tgg gct cag ctc aca ttc tgg 417
 Leu Cys Ile Leu Ala Ser Gln His Ala Trp Ala Gln Leu Thr Phe Trp
 45 50 55 60
 gag gcc agc cag ctt tac ctg ctg ttc ctg agc ctt acg ctg gcc act 465
 Glu Ala Ser Gln Leu Tyr Leu Leu Phe Leu Ser Leu Thr Leu Ala Thr
 65 70 75
 25 gtc aac gcc cgc tgg ctg gaa ccc cgc acc aca gct gcc atg tgg gcc 513
 Val Asn Ala Arg Trp Leu Glu Pro Arg Thr Thr Ala Ala Met Trp Ala
 80 85 90
 ctg caa acc gtg gag aag gag cga ggc ctg ggt ggg gag gta cca ggc 561
 Leu Gln Thr Val Glu Lys Glu Arg Gly Leu Gly Gly Glu Val Pro Gly
 30 95 100 105
 agc cac cag ggt ccc gat ccc tac cgc cag ctg cga gag aag gac ccc 609
 Ser His Gln Gly Pro Asp Pro Tyr Arg Gln Leu Arg Glu Lys Asp Pro
 110 115 120
 aag tac agt gct ctc cgc cag aat ttc ttc cgc tac cat ggg ctg tcc 657
 35 Lys Tyr Ser Ala Leu Arg Gln Asn Phe Phe Arg Tyr His Gly Leu Ser
 125 130 135 140
 tct ctt tgc aat ctg ggc tgc gtc ctg agc aat ggg ctc tgt ctc gct 705
 Ser Leu Cys Asn Leu Gly Cys Val Leu Ser Asn Gly Leu Cys Leu Ala
 145 150 155
 40 ggc ctt gcc ctg gaa ata agg agc ctc tagcatgggc cctgcatgct 752
 Gly Leu Ala Leu Glu Ile Arg Ser Leu
 160 165
 aataaatgct tctccaaaaa aaaaaaaaaa aa 784

 45 <210> 116
 <211> 804
 <212> DNA
 <213> Homo sapiens

 50 <220>
 <221> CDS
 <222> 75..623

 <220>
 55 <221> sig_peptide
 <222> 75..215
 <223> Von Heijne matrix
 score 8.34104221735598
 seq RLLLPCLVRMALC/AP

 60 <400> 116
 agtacggtgg ccgacgggag tcagacgctg gggatgaatg aaggtgctgg gtgcaggatc 60
 aacaaacagt aata atg act gaa tgt aca agt ctt cag ttt gtc agc cct 110

					Met	Thr	Glu	Cys	Thr	Ser	Leu	Gln	Phe	Val	Ser	Pro		
							-45					-40						
	ttt	gct	ttt	gag	gca	atg	cag	aag	gtg	gat	gtt	gtt	tgc	ctg	gca	tct	158	
	Phe	Ala	Phe	Glu	Ala	Met	Gln	Lys	Val	Asp	Val	Val	Cys	Leu	Ala	Ser		
5	-35					-30					-25					-20		
	tta	agt	gat	cca	gaa	tta	aga	ctt	ctt	ctg	ccc	tgt	ttg	gta	cgg	atg	206	
	Leu	Ser	Asp	Pro	Glu	Leu	Arg	Leu	Leu	Leu	Pro	Cys	Leu	Val	Arg	Met		
					-15						-10					-5		
	gca	ctt	tgt	gca	cct	gct	gac	cag	agc	caa	agc	tgg	gct	cag	gat	aag	254	
10	Ala	Leu	Cys	Ala	Pro	Ala	Asp	Gln	Ser	Gln	Ser	Trp	Ala	Gln	Asp	Lys		
				1				5					10					
	aaa	ctc	atc	ctt	cgc	ctt	ctt	tct	gga	gtg	gaa	gct	gtc	aac	tcc	att	302	
	Lys	Leu	Ile	Leu	Arg	Leu	Leu	Ser	Gly	Val	Glu	Ala	Val	Asn	Ser	Ile		
		15					20					25						
15	ggt	gca	ttg	ttg	tcc	gtg	gac	ttt	cat	gct	tta	gaa	caa	gat	gcc	agc	350	
	Val	Ala	Leu	Leu	Ser	Val	Asp	Phe	His	Ala	Leu	Glu	Gln	Asp	Ala	Ser		
	30					35					40				45			
	aaa	gaa	cag	cag	ctt	aga	ccg	agt	ctt	gcc	ctg	ttg	ccc	agg	ctg	gag	398	
	Lys	Glu	Gln	Gln	Leu	Arg	Pro	Ser	Leu	Ala	Leu	Leu	Pro	Arg	Leu	Glu		
20					50					55					60			
	tgc	ggt	ggc	gtg	atc	tcg	gct	cac	tgc	aac	ctc	cac	ctc	ctg	ggt	tca	446	
	Cys	Gly	Gly	Val	Ile	Ser	Ala	His	Cys	Asn	Leu	His	Leu	Leu	Gly	Ser		
				65					70				75					
	agt	gat	tct	tct	gcc	tca	gtc	tcc	cga	gta	gat	ggg	act	aca	ggc	acg	494	
25	Ser	Asp	Ser	Ser	Ala	Ser	Val	Ser	Arg	Val	Asp	Gly	Thr	Thr	Gly	Thr		
		80						85					90					
	cgc	cac	cat	gcc	cgg	ctt	ttt	tgt	att	att	agt	aga	gac	gag	gtt	tca	542	
	Arg	His	His	Ala	Arg	Leu	Phe	Cys	Ile	Ile	Ser	Arg	Asp	Glu	Val	Ser		
		95					100					105						
30	cca	tat	tgg	cca	ggc	tgg	tct	cga	act	ccc	aac	ctt	gtg	atc	cac	ctg	590	
	Pro	Tyr	Trp	Pro	Gly	Trp	Ser	Arg	Thr	Pro	Asn	Leu	Val	Ile	His	Leu		
	110					115					120				125			
	cct	cag	cct	ccc	aaa	gta	ctg	gga	tta	ccg	gcg	tgagccactg	tgccctggcct				643	
	Pro	Gln	Pro	Pro	Lys	Val	Leu	Gly	Leu	Pro	Ala							
35					130				135									
	atgtggtgga	gtatttatta	tacgtaggat	gtgaatccct				gaaatacacaca	ggcaaactaa									703
	atagcatttc	agaagtaaca	gaacatttta	gaacacttta				tacatccttt	tatagcttat									763
	ttcaataaaaa	gataattttt	atacaaaaaa	aaaaaaaaaa				a									804	
40	<210> 117																	
	<211> 484																	
	<212> DNA																	
	<213> Homo sapiens																	
45	<220>																	
	<221> CDS																	
	<222> 30..335																	
	<220>																	
50	<221> sig_peptide																	
	<222> 30..71																	
	<223> Von Heijne matrix																	
	score 4.49063834776683																	
	seq FLTALLWRGRIPG/RQ																	
55	<400> 117																	
	gcagagtctt	gagcagcgcg	gcaggcacc	atg	ttc	ctg	act	gcg	ctc	ctc	tgg						53	
				Met	Phe	Leu	Thr	Ala	Leu	Leu	Trp							
							-10											
60	cgc	ggc	cgc	att	ccc	ggc	cgt	cag	tgg	atc	ggg	aag	cac	cgg	cgg	ccg	101	
	Arg	Gly	Arg	Ile	Pro	Gly	Arg	Gln	Trp	Ile	Gly	Lys	His	Arg	Arg	Pro		
	-5					1				5					10			
	cgg	ttc	gtg	tcg	ttg	cgc	gcc	aag	cag	aac	atg	atc	cgc	cgc	ctg	gag	149	


```

Arg Phe Val Ser Leu Arg Ala Lys Gln Asn Met Ile Arg Arg Leu Glu
      15      20      25
atc gag gcg gag aac cat tac tgg ctg agc atg ccc tac atg acc cgg      197
Ile Glu Ala Glu Asn His Tyr Trp Leu Ser Met Pro Tyr Met Thr Arg
5      30      35      40
gag cag gag cgc ggc cac gcc gcg gtg cgc agg agg gag gcc ttc gag      245
Glu Gln Glu Arg Gly His Ala Ala Val Arg Arg Arg Glu Ala Phe Glu
      45      50      55
gcc ata aag gcg gcc gcc act tcc aag ttc ccc ccg cat aga ttc att      293
10 Ala Ile Lys Ala Ala Ala Thr Ser Lys Phe Pro Pro His Arg Phe Ile
      60      65      70
gcg gac cag ctc gac cat ctc aat gtc acc aag aaa tgg tcc      335
Ala Asp Gln Leu Asp His Leu Asn Val Thr Lys Lys Trp Ser
      75      80      85
15 taatcctgag tcgtcaccct tggattttat ggatcacgga gctgaccatc tttacctggt      395
cctggaactg aaaaactgta gcttgtgtga aaatgagcct ttggaccagt ctttattaaa      455
acaaacaaac acaaaaaaaaa aaaaaaaaaa      484

<210> 118
20 <211> 985
    <212> DNA
    <213> Homo sapiens

    <220>
25 <221> CDS
    <222> 21..752

    <220>
    <221> sig_peptide
30 <222> 21..107
    <223> Von Heijne matrix
        score 3.61056351168286
        seq FPLYLLNFLGLWS/WI

35 <400> 118
gtttttttcc cttctgagca atg gag ctt acc atc ttt atc ctg aga ctg gcc      53
      Met Glu Leu Thr Ile Phe Ile Leu Arg Leu Ala
      -25      -20
att tac atc ctg aca ttt ccc ttg tac ctg ctg aac ttt ctg ggc ttg      101
40 Ile Tyr Ile Leu Thr Phe Pro Leu Tyr Leu Leu Asn Phe Leu Gly Leu
      -15      -10      -5
tgg agc tgg ata tgc aaa aaa tgg ttc ccc tac ttc ttg gtg agg ttc      149
Trp Ser Trp Ile Cys Lys Lys Trp Phe Pro Tyr Phe Leu Val Arg Phe
      1      5      10
45 act gtg ata tac aac gaa cag atg gca agc aag cgg gag ctc ttc      197
Thr Val Ile Tyr Asn Glu Gln Met Ala Ser Lys Lys Arg Glu Leu Phe
      15      20      25      30
agt aac ctg cag gag ttt gcg ggc ccc tcc ggg aaa ctc tcc ctg ctg      245
Ser Asn Leu Gln Glu Phe Ala Gly Pro Ser Gly Lys Leu Ser Leu Leu
      35      40      45
50 gaa gtg ggc tgt ggc acg ggg gcc aac ttc aag ttc tac cca cct ggg      293
Glu Val Gly Cys Gly Thr Gly Ala Asn Phe Lys Phe Tyr Pro Pro Gly
      50      55      60
tgc agg gtg acc tgt att gac ccc aac ccc aac ttt gag aag ttt ttg      341
55 Cys Arg Val Thr Cys Ile Asp Pro Asn Pro Asn Phe Glu Lys Phe Leu
      65      70      75
atc aag agc att gca gag aac cga cac ctg cag ttt gag cgc ttt gtg      389
Ile Lys Ser Ile Ala Glu Asn Arg His Leu Gln Phe Glu Arg Phe Val
      80      85      90
60 gta gct gcc ggg gag aac atg cac cag gtg gct gat ggc tct gtg gat      437
Val Ala Ala Gly Glu Asn Met His Gln Val Ala Asp Gly Ser Val Asp
      95      100      105      110
gtg gtg gtc tgc acc ctg gtg ctg tgc tct gtg aag aac cag gag cgg      485

```

	Val	Val	Val	Cys	Thr	Leu	Val	Leu	Cys	Ser	Val	Lys	Asn	Gln	Glu	Arg	
					115					120					125		
	att	ctc	cgc	gag	gtg	tgc	aga	gtg	ctg	aga	ccg	gga	ggg	gct	ttc	tat	533
	Ile	Leu	Arg	Glu	Val	Cys	Arg	Val	Leu	Arg	Pro	Gly	Gly	Ala	Phe	Tyr	
5				130					135					140			
	ttc	atg	gag	cat	gtg	gca	gct	gag	tgt	tcg	act	tgg	aat	tac	ttc	tgg	581
	Phe	Met	Glu	His	Val	Ala	Ala	Glu	Cys	Ser	Thr	Trp	Asn	Tyr	Phe	Trp	
				145					150					155			
	caa	caa	gtc	ctg	gat	cct	gcc	tgg	cac	ctt	ctg	ttt	gat	ggg	tgc	aac	629
10	Gln	Gln	Val	Leu	Asp	Pro	Ala	Trp	His	Leu	Leu	Phe	Asp	Gly	Cys	Asn	
				160					165					170			
	ctg	acc	aga	gag	agc	tgg	aag	gcc	ctg	gag	cgg	gcc	agc	ttc	tct	aag	677
	Leu	Thr	Arg	Glu	Ser	Trp	Lys	Ala	Leu	Glu	Arg	Ala	Ser	Phe	Ser	Lys	
				175			180					185				190	
15	ctg	aag	ctg	cag	cac	atc	cag	gcc	cca	ctg	tcc	tgg	gag	ttg	gtg	cgc	725
	Leu	Lys	Leu	Gln	His	Ile	Gln	Ala	Pro	Leu	Ser	Trp	Glu	Leu	Val	Arg	
				195					200					205			
	cct	cat	atc	tat	gga	tat	gct	gtg	aaa	tagtgtgagc	tggcagttaa						772
	Pro	His	Ile	Tyr	Gly	Tyr	Ala	Val	Lys								
20				210					215								
	gagctgaatg	gctcaaagaa	tttaaagctt	cagttttaca	tttaaaatgc	taggtgggtg											832
	cctgtaatcc	caggtacttg	gaaggctgag	gcaggagaat	ctcttgaacc	cagaaggcga											892
	aggttgacgt	gaaccgagat	catgccattg	tactctagcc	tgggtgacaa	gagcaagact											952
	ccgtctcaaa	aaaaaataaa	aaaaaaaaaa	aaa													985
25																	
	<210>	119															
	<211>	839															
	<212>	DNA															
	<213>	Homo sapiens															
30																	
	<220>																
	<221>	CDS															
	<222>	185..715															
35																	
	<220>																
	<221>	sig_peptide															
	<222>	185..253															
	<223>	Von Heijne matrix															
		score 9.49395175807817															
40		seq SLLFICFFGESFC/IC															
	<400>	119															
	atatttttgc	tactggcaag	gttatatgaa	gtgcttttat	tgaagcacca	ttttaactaa											60
	cagctcctgg	tattttctgc	ttcccttcgt	aggggaattta	gttattttat	tttattattt											120
45	agctaattta	gctattttta	aatagctaaa	ttttagctac	ttttttttca	attgacaaaag											180
	aagg atg tct	aat caa aga	cta ccg ctg	att ttt tct	ctg ttg ttt	atc											229
		Met Ser Asn	Gln Arg Leu	Pro Leu Ile	Phe Ser Leu	Leu Phe Ile											
			-20		-15	-10											
	tgc ttc ttc	ggg gag agt	ttc tgc att	tgt gat gga	act gtc tgg	aca											277
50	Cys Phe Phe	Gly Glu Ser	Phe Cys Ile	Cys Asp Gly	Thr Val Trp	Thr											
			-5		1	5											
	aag gtt gga	tgg gag att	ctt cca gaa	gaa gta cat	tat tgg aaa	ggt											325
	Lys Val Gly	Trp Glu Ile	Leu Pro Glu	Glu Val His	Tyr Trp Lys	Val											
			10		15	20											
55	aag ggt tct	cca tct cac	tgc ctg cct	tat ctt ctg	gat aaa cta	tgc											373
	Lys Gly Ser	Pro Ser His	Cys Leu Pro	Tyr Leu Leu	Asp Lys Leu	Cys											
			25		30	35											40
	tgc gac ttt	gct aac atg	gat ata ttt	cag ggt tgt	tta tat ctc	att											421
	Cys Asp Phe	Ala Asn Met	Asp Ile Phe	Gln Gly Cys	Leu Tyr Leu	Ile											
60			45		50	55											
	tat aat tta	tta caa gct	gtc ttc ttc	gtc tta ttt	ggt ttg tct	gtg											469
	Tyr Asn Leu	Leu Gln Ala	Val Phe Phe	Val Leu Phe	Val Leu Ser	Val											
			60		65	70											

```

cat tac ctg tgg aag aaa tgg aag aaa cac caa aaa aag ctg aaa aag      517
His Tyr Leu Trp Lys Lys Trp Lys Lys His Gln Lys Lys Leu Lys Lys
      75                      80                      85
caa gcc tcc tta gaa aaa cct ggt aat gat cta gaa agc cca ttg atc      565
5 Gln Ala Ser Leu Glu Lys Pro Gly Asn Asp Leu Glu Ser Pro Leu Ile
      90                      95                      100
aac aac att gac caa aca ctc cac aga gtg gca acc aca gca tca gtg      613
Asn Asn Ile Asp Gln Thr Leu His Arg Val Ala Thr Thr Ala Ser Val
      105                      110                      115                      120
10 ata tac aag atc tgg gag cac agg tct cac cat cct tcc tct aag aaa      661
Ile Tyr Lys Ile Trp Glu His Arg Ser His His Pro Ser Ser Lys Lys
      125                      130                      135
att aag cac tgc aaa tta aag aag aag agt aaa gaa gaa gga gcc aga      709
Ile Lys His Cys Lys Leu Lys Lys Lys Ser Lys Glu Glu Gly Ala Arg
      140                      145                      150
15 aga tac taaataaatg catatgcaaa tgtagcttac tcaattatag atatcacaaa      765
Arg Tyr
agaaatctat catctaagga ttaaaaattg ttctttggaa acctttataa aaaaaaaaga      825
aaaaaaaaaa aaaa      839
20
<210> 120
<211> 583
<212> DNA
<213> Homo sapiens
25
<220>
<221> CDS
<222> 54..527
30
<220>
<221> sig_peptide
<222> 54..116
<223> Von Heijne matrix
      score 6.80928714315144
35      seq ALXSLNLAPPTVA/AP

<400> 120
aacgtcatct aggagcaccg agcagcttgg ctaaaagtaa ggggtgtcgtg ctg atg      56
Met
40 gcc ctg tgc gca ctg acc cgc gct ctg ccs tct ctg aac ctg gcg ccc      104
Ala Leu Cys Ala Leu Thr Arg Ala Leu Pro Ser Leu Asn Leu Ala Pro
-20                      -15                      -10                      -5
ccg acc gtc gcc gcc cct gcc ccg agt ctg ttc ccc gcc gcc cag atg      152
Pro Thr Val Ala Ala Pro Ala Pro Ser Leu Phe Pro Ala Ala Gln Met
45      1                      5                      10
atg aac aat ggc ctc ctc caa cag ccc tct gcc ttg atg ttg ctc ccc      200
Met Asn Asn Gly Leu Leu Gln Gln Pro Ser Ala Leu Met Leu Leu Pro
      15                      20                      25
tgc cgc cca gtt ctt act tct gtg gcc ctt aat gcc aac ttt gtg tcc      248
50 Cys Arg Pro Val Leu Thr Ser Val Ala Leu Asn Ala Asn Phe Val Ser
      30                      35                      40
tgg aag agt cgt acc aag tac acc att aca cca gtg aag atg agg aag      296
Trp Lys Ser Arg Thr Lys Tyr Thr Ile Thr Pro Val Lys Met Arg Lys
      45                      50                      55                      60
55 tct ggg ggc cga gac cac aca ggt gct gga aac gtg cgt aga aca gta      344
Ser Gly Gly Arg Asp His Thr Gly Ala Gly Asn Val Arg Arg Thr Val
      65                      70                      75
ggc cga gta tcc aac gtt gat cat aac aaa cgg gtc att ggc aag gca      392
Gly Arg Val Ser Asn Val Asp His Asn Lys Arg Val Ile Gly Lys Ala
      80                      85                      90
60 ggt cgc aac cgc tgg ctg ggc aag agg cct aac agt ggg cgg tgg cac      440
Gly Arg Asn Arg Trp Leu Gly Lys Arg Pro Asn Ser Gly Arg Trp His
      95                      100                      105

```

	cgc aag ggg ggc tgg gct ggc cga aag att cgg cca cta ccc ccc atg	488
	Arg Lys Gly Gly Trp Ala Gly Arg Lys Ile Arg Pro Leu Pro Pro Met	
	110 115 120	
	aag agt tac gtg aag ctg cct tct gct tct gcc caa agc tgatatccct	537
5	Lys Ser Tyr Val Lys Leu Pro Ser Ala Ser Ala Gln Ser	
	125 130 135	
	gtactctaataaaaatgcccccctcaaaaaaaaaaaaaaa	583
	<210> 121	
10	<211> 1024	
	<212> DNA	
	<213> Homo sapiens	
	<220>	
15	<221> CDS	
	<222> 129..686	
	<220>	
	<221> sig_peptide	
20	<222> 129..185	
	<223> Von Heijne matrix	
	score 6.45239823575329	
	seq SVFLLMVNGQVES/AQ	
25	<400> 121	
	cttcgcgaag gtgtcgctgc caagaaacgt gtccctgcgcg ctacgccgctc tgtttctagg	60
	gcaacgcccg cgtctcttag caaccgcgcg cggcctagggt gggccccccc ggcacccccca	120
	gacctgcc atg gcg acc gcg agt cct agc gtc ttt cta ctc atg gtc aac	170
	Met Ala Thr Ala Ser Pro Ser Val Phe Leu Leu Met Val Asn	
30	-15 -10	
	ggg cag gtg gag agc gcc cag ttt cca gag tat gat gac ctc tac tgc	218
	Gly Gln Val Glu Ser Ala Gln Phe Pro Glu Tyr Asp Asp Leu Tyr Cys	
	-5 1 5 10	
	aag tac tgc ttt gtg tac ggc cag gac tgg gcc ccc aca gcg ggt ctg	266
35	Lys Tyr Cys Phe Val Tyr Gly Gln Asp Trp Ala Pro Thr Ala Gly Leu	
	15 20 25	
	gag gag ggg atc tca cag atc aca tcc aag agc caa gat gtg cgg caa	314
	Glu Glu Gly Ile Ser Gln Ile Thr Ser Lys Ser Gln Asp Val Arg Gln	
	30 35 40	
40	gca ctg gtg tgg aac ttc ccc att gat gtc acc ttt aaa agc acc aac	362
	Ala Leu Val Trp Asn Phe Pro Ile Asp Val Thr Phe Lys Ser Thr Asn	
	45 50 55	
	ccc tac ggc tgg cca cag atc gtg ctc agc gtg tat gga cca gat gtg	410
	Pro Tyr Gly Trp Pro Gln Ile Val Leu Ser Val Tyr Gly Pro Asp Val	
45	60 65 70 75	
	ttc ggg aac gat gtg gtt cga ggc tat ggg gcc gtg cac gtg ccc ttc	458
	Phe Gly Asn Asp Val Val Arg Gly Tyr Gly Ala Val His Val Pro Phe	
	80 85 90	
	tca cct ggc cgg cac aaa agg acc atc ccc atg ttt gtc cca gaa tct	506
50	Ser Pro Gly Arg His Lys Arg Thr Ile Pro Met Phe Val Pro Glu Ser	
	95 100 105	
	acg tct aaa ctg cag aag ttt aca agc tgg ttc atg ggg cgg cgg ccc	554
	Thr Ser Lys Leu Gln Lys Phe Thr Ser Trp Phe Met Gly Arg Arg Pro	
	110 115 120	
55	gag tac aca gac ccc aag gtg gct cag ggt gaa ggc cgg gaa gct	602
	Glu Tyr Thr Asp Pro Lys Val Val Ala Gln Gly Glu Gly Arg Glu Ala	
	125 130 135	
	atc aca gct ccc cgg aaa gct gtc ttc tct gtc cat ggc ctc acc tca	650
	Ile Thr Ala Pro Arg Lys Ala Val Phe Ser Val His Gly Leu Thr Ser	
60	140 145 150 155	
	ccc agg gca ctg gcc ttg gtc cac atc aag ggg acc tgaagcttcc	696
	Pro Arg Ala Leu Ala Leu Val His Ile Lys Gly Thr	
	160 165	

```

ctgaagcctc tagcctgtgg tgtgcacgta caagcctcag gccccatttg tccagcctgt 756
cagcagctgg gaaatactaa gtcaccctct tctgggtatg ttttaatttc caatttttct 816
caacattact gaaatgtcta aatgtggaaa agttgacatc attttacagt gaacaccaca 876
taccaccac ctagatttta ccattaccaa tttcctgttc cgtacttgta tattcacata 936
5 tatccaacta ttcacccctg cttcaatcca tcctattttt attgcatttc aaaataaact 996
gtgaaatcag gaaaaaaaaa aaaaaaaaa 1024

<210> 122
<211> 760
10 <212> DNA
    <213> Homo sapiens

<220>
<221> CDS
15 <222> 165..614

<220>
<221> sig_peptide
<222> 165..305
20 <223> Von Heijne matrix
    score 5.10820788278539
    seq ALGLALCSTKALS/VG

<400> 122
25 aatttccgat gccaggcacc ctcaaggcac agaggctggg gctcatgttg ggggcacttg 60
   gcctctccag gcctcgaagg cttctcctgg gctgatgcga gctgggggaac gggagggacg 120
   gacgtgggag cgagaacgtc acactggagg cagctgggtg cacg atg ggg gac aga 176
                                     Met Gly Asp Arg
                                     -45
30 gtg aaa ggt agc aag tca aga gcc ttc gtg tca cca tgg cca cac acc 224
   Val Lys Gly Ser Lys Ser Arg Ala Phe Val Ser Pro Trp Pro His Thr
                                     -40 -35 -30
   ccg atg gct tcc ggc ttg agg gac ccc tgg ctg cag ccc aca gcc ctg 272
   Pro Met Ala Ser Gly Leu Arg Asp Pro Trp Leu Gln Pro Thr Ala Leu
   -25 -20 -15
35 ggc ctt gca ctg tgc tct acg aag gcc ctg tcc gtg ggc tct gcc cct 320
   Gly Leu Ala Leu Cys Ser Thr Lys Ala Leu Ser Val Gly Ser Ala Pro
   -10 -5 1 5
   ttg ccc ccg cga aat tcc aac acc atg gcg gcg gct gcc ctg gct gcc 368
   Leu Pro Pro Arg Asn Ser Asn Thr Met Ala Ala Ala Ala Leu Ala Ala
   10 15 20
   ccc agc ctg ggc ttc gat ggg gtg att ggg gtg ctc gtg gct gat acc 416
   Pro Ser Leu Gly Phe Asp Gly Val Ile Gly Val Leu Val Ala Asp Thr
   25 30 35
45 agc ctc acg gac atg cac gtg gtg gat gta gag ctg agc gga ccc cgg 464
   Ser Leu Thr Asp Met His Val Val Asp Val Glu Leu Ser Gly Pro Arg
   40 45 50
   ggc ccc acg ggc cga agc ttt gct gtg cac acc cgc aga gag aac cct 512
   Gly Pro Thr Gly Arg Ser Phe Ala Val His Thr Arg Arg Glu Asn Pro
   55 60 65
   gcc gag cca ggc gcg gtc acc ggc tcc gcc acc gtc acg gcc ttc tgg 560
   Ala Glu Pro Gly Ala Val Thr Gly Ser Ala Thr Val Thr Ala Phe Trp
   70 75 80 85
   cgg agc ctc ctg gcc tgc tgc cag ctc ccc tcc agg ccg ggg atc cat 608
   Arg Ser Leu Leu Ala Cys Cys Gln Leu Pro Ser Arg Pro Gly Ile His
   90 95 100
   ctc tgc tgagaagcct cctccctccc gagacaagat catctgcctg gcctctcacc 664
   Leu Cys
   accaccatcc caccctgcc ctgccccact tccccagggt ctcccttctg actcagtaaa 724
60 gatcaccgct gcctccctca aaaaaaaaaa aaaaaa 760

<210> 123
<211> 594

```

<212> DNA
<213> Homo sapiens

<220>
5 <221> CDS
<222> 192..476

<220>
<221> sig_peptide
10 <222> 192..326
<223> Von Heijne matrix
score 6.60884760057354
seq FILLLLLSGPAEM/SA

15 <400> 123
actttttattg aaaaagacta cagcaaatca tactgaggtg aatgaagaca gtgaaatgaa 60
ggagaaggca ggtcctcttt atgttttcgc agctgggtca aggggtttgg gggttttctat 120
ctagggttaaa gattgcgtaa tacacagctg gagccataga cattaatgca tggtttatcac 180
acgcaacaac g atg ctg cat cat gtg att aca gct ggg cct gtg ctg ctt 230
20 Met Leu His His Val Ile Thr Ala Gly Pro Val Leu Leu
-45 -40 -35
cta cac ctg cct cgc cct gac act tcc acc agg ttg ctg ctg acc tcc 278
Leu His Leu Pro Arg Pro Asp Thr Ser Thr Arg Leu Leu Leu Thr Ser
-30 -25 -20
25 gtc tct gct ttt atc ctg tta ctg ctg ctt tca gga cca gca gaa atg 326
Val Ser Ala Phe Ile Leu Leu Leu Leu Leu Ser Gly Pro Ala Glu Met
-15 -10 -5
tca gct tcc cag gaa tcc ttc cct gga tct ctg cag caa gaa ata gct 374
Ser Ala Ser Gln Glu Ser Phe Pro Gly Ser Leu Gln Gln Glu Ile Ala
30 1 5 10 15
tct ctg atc act gta gca ctt ggt tct tta ata tct tta tct tgc tct 422
Ser Leu Ile Thr Val Ala Leu Gly Ser Leu Ile Ser Leu Ser Cys Ser
20 25 30
acc ttg tta tat ttt tct tgt gaa ctt aaa att ccc tgt gag gac gta 470
35 Thr Leu Leu Tyr Phe Ser Cys Glu Leu Lys Ile Pro Cys Glu Asp Val
35 40 45
aac ctt tgaaggtatg tctcatatct ctgaacctct ttaaaatgcc tagcatccct 526
Asn Leu
50
40 gtgtgggtgc caattgcttg tgtattgaat taaattgtga ttgttaactt gaaaaaaaaa 586
aaaaaaaaa 594

<210> 124
<211> 559
45 <212> DNA
<213> Homo sapiens

<220>
<221> CDS
50 <222> 16..297

<220>
<221> sig_peptide
<222> 16..93
55 <223> Von Heijne matrix
score 6.65836819891491
seq FCGSACLLAVIRA/FF

<400> 124
60 ttacacaggg gataa atg gca gca atc gag att gaa gtc aag cct aac cag 51
Met Ala Ala Ile Glu Ile Glu Val Lys Pro Asn Gln
-25 -20 -15
ggc ttt tgc ggg agc gca tgc ctt ttg gct gta att cgt gca ttt ttt 99

Gly Phe Cys Gly Ser Ala Cys Leu Leu Ala Val Ile Arg Ala Phe Phe
-10 -5 1
147
5 ttt aag aaa aac gcc tgc ctt ctg cgt gag att ctc cag agc aaa ctg
Phe Lys Lys Asn Ala Cys Leu Leu Arg Glu Ile Leu Gln Ser Lys Leu
5 5 10 15
195
ggc ggc atg ggc cct gtg gtc ttt tcg tac aga ggg ctt cct ctt tgg
Gly Gly Met Gly Pro Val Val Phe Ser Tyr Arg Gly Leu Pro Leu Trp
20 25 30
243
10 ctc ttt gcc tgg ttg ttt cca aga tgt act gtg cct ctt act ttc ggt
Leu Phe Ala Trp Leu Phe Pro Arg Cys Thr Val Pro Leu Thr Phe Gly
35 40 45 50
291
ttt gaa aac atg agg ggg ttg ggc gtg gta gct tac gcc tgt aat ccc
Phe Glu Asn Met Arg Gly Leu Gly Val Val Ala Tyr Ala Cys Asn Pro
55 60 65
347
15 agc act tagggaggcc gaggcgggag gatggcttga ggtccgtagt tgagaccagc
Ser Thr
407
ctggccaaca tggatgaagcc tgggtctctac aaaaaaataa taacaaaaat tagccgggtg
tggatggctcg tgccctgtggt cccagctgct ccggtggctg aggcgggagg atctcttgag
467
cttaggcttt tgagctatca tggcgccagt gcactccagc gtgggcaaca gagcgagacc
527
20 ctgtctctca aaaacaaaaa aaaaaaaaaa aa 559

<210> 125
<211> 744
<212> DNA
25 <213> Homo sapiens

<220>
<221> CDS
<222> 216..635
30
<220>
<221> sig_peptide
<222> 216..335
<223> Von Heijne matrix
35 score 4.38054120608596
seq ITLVSAAPGKVIC/EM

<400> 125
gcttcgtcac aagggtgcga tgaaagtcag tgagcaaatac ggcggaccacc ggggctgccca 60
40 gctcgccctga ctcccgccct cttgcgctcc tagggggcgga gaagggtgcg ggctcttcgc 120
cctttgtgtc ctccctcttt cactaacttc tggactttcc agctcttcg aagttcgttc 180
ttgcgcaaag cccaaaggct ggaaaaccgt ccacg atg acc agc atg act cag 233
Met Thr Ser Met Thr Gln
-40 -35
45 tct ctg cgg gag gtg ata aag gcc atg acc aag gct cgc aat ttt gag 281
Ser Leu Arg Glu Val Ile Lys Ala Met Thr Lys Ala Arg Asn Phe Glu
-30 -25 -20
329
aga gtt ttg gga aag att act ctt gtc tct gct gct cct ggg aaa gtg
Arg Val Leu Gly Lys Ile Thr Leu Val Ser Ala Ala Pro Gly Lys Val
50 -15 -10 -5
377
att tgt gaa atg aaa gta gaa gaa gag cat acc aat gca ata ggc act
Ile Cys Glu Met Lys Val Glu Glu Glu His Thr Asn Ala Ile Gly Thr
1 5 10
425
55 ctc cac ggc ggt ttg aca gcc acg tta gta gat aac ata tca aca atg
Leu His Gly Gly Leu Thr Ala Thr Leu Val Asp Asn Ile Ser Thr Met
15 20 25 30
473
gct ctg cta tgc acg gaa agg gga gca ccc gga gtc agt gtc gat atg
Ala Leu Leu Cys Thr Glu Arg Gly Ala Pro Gly Val Ser Val Asp Met
35 40 45
521
60 aac ata acg tac atg tca cct gca aaa tta gga gaa gat ata gtg att
Asn Ile Thr Tyr Met Ser Pro Ala Lys Leu Gly Glu Asp Ile Val Ile
50 55 60
569
aca gca cat gtt ctg aag caa gga aaa aca ctt gca ttt acc tct gtg

Thr Ala His Val Leu Lys Gln Gly Lys Thr Leu Ala Phe Thr Ser Val
 65 70 75
 gat ctg acc aac aag gcc aca gga aaa tta ata gca caa gga aga cac 617
 Asp Leu Thr Asn Lys Ala Thr Gly Lys Leu Ile Ala Gln Gly Arg His
 5 80 85 90
 aca aaa cac ctg gga aac tgagagaaca gcagaatgac ctaaagaaac 665
 Thr Lys His Leu Gly Asn
 95 100
 ccaacaatga atatcaagta tagatttgac tcaaacaatt gtaatttttg aaataaacta 725
 10 gcaaaaaaaaa aaaaaaaaaa 744

 <210> 126
 <211> 824
 <212> DNA
 15 <213> Homo sapiens

 <220>
 <221> CDS
 <222> 164..280
 20
 <220>
 <221> sig_peptide
 <222> 164..268
 <223> Von Heijne matrix
 25 score 5.73290676305402
 seq TLPLCPVTSPVWG/WS

 <400> 126
 30 tgtgttcaat cgtgtgaatg gccggcgggc cccctccacg tccccatcct tcgaggggac 60
 ccaggagacc tacacagtgg ccacagagga gaatgtccgc tttgtgtccg aaggtagcga 120
 gcggggccag aggggtgcggc ataggctgct gggtcgcaaa acc atg gac ccg gga 175
 Met Asp Pro Gly
 -35
 tgg ccc cac ttc aag ctg acc cac agc cgc tgc atg gct gtg ctt ttc 223
 35 Trp Pro His Phe Lys Leu Thr His Ser Arg Cys Met Ala Val Leu Phe
 -30 -25 -20
 ctt ggc act ctg ccc ttg tgt cct gtg acc agc cct gtg tgg ggc tgg 271
 Leu Gly Thr Leu Pro Leu Cys Pro Val Thr Ser Pro Val Trp Gly Trp
 -15 -10 -5 1
 40 agt cca ggg tgaccatcag gccctgggtg ggcgatgggg tgccctgggac 320
 Ser Pro Gly
 ctggctcagc ccgactgccc tccctcccaca gcctggcagc aggtgcaaca gcagctggat 380
 ggtggcccgag ccggtgaggg cgggccaagg cctgtgcagt acgtggagag gacccccaat 440
 ccccggtctg agaactttgt gcccatattac ctagacgagt ggtgggcgca gcagttcctg 500
 45 gcgagaatca ccagctgttc ctagtggtgctg ctgggagggg gcgctgctac acggccgacc 560
 tgtcgccagg agagaagcat ggcgccctgc ccacccactg cgccctggctg ggtgccggcc 620
 acacctgaag tgccagcatt tggacttttg cacctttttt tcccttggcc cggctgtccc 680
 aaccaagctg ccatggccaa gggccgaacc cgtctgacct cagccctgct cactgtgccc 740
 agggaccagc gaccagcccc tggggctggc agggaggagc tccaggctaa taaagtggag 800
 50 aaactgtcaa aaaaaaaaaa aaaa 824

 <210> 127
 <211> 526
 <212> DNA
 55 <213> Homo sapiens

 <220>
 <221> CDS
 <222> 68..301
 60
 <220>
 <221> sig_peptide
 <222> 68..190

<223> Von Heijne matrix
score 4.68908216483476
seq AYLLYILLTGALQ/FG

5 <400> 127
acatccggtg tggctcgacgg gtcctccaag agtttggggc gcggaccgga gtaccttgcg 60
tgcagtt atg tcg gcg tcg gta gtg tct gtc att tcg cgg ttc tta gaa 109
Met Ser Ala Ser Val Val Ser Val Ile Ser Arg Phe Leu Glu
-40 -35 -30
10 gag tac ttg agc tcc act ccg cag cgt ctg aag ttg ctg gac gcg tac 157
Glu Tyr Leu Ser Ser Thr Pro Gln Arg Leu Lys Leu Leu Asp Ala Tyr
-25 -20 -15
ctg ctg tat ata ctg ctg acc ggg gcg ctg cag ttc ggt tac tgt ctc 205
Leu Leu Tyr Ile Leu Leu Thr Gly Ala Leu Gln Phe Gly Tyr Cys Leu
-10 -5 1 5
15 ctc gtg ggg acc ttc ccc ttc aac tct ttt ctc tcg ggc ttc atc tct 253
Leu Val Gly Thr Phe Pro Phe Asn Ser Phe Leu Ser Gly Phe Ile Ser
10 15 20
tgt gtg ggg agt ttc atc cta gcg ggt tca ctc ttt gaa ttt cct gga 301
20 Cys Val Gly Ser Phe Ile Leu Ala Gly Ser Leu Phe Glu Phe Pro Gly
25 30 35
taagagttct ggagatggca gcttattgga cacatggatt ttcttcagat ttgcacttac 361
tgctagctct gctttttatg caggagaaaa gccagagatt cactgtgtgt cagaacaact 421
ttctaacaaa catttattaa tccagcctct gcctttcatt aaatgtaacc ttttgccttc 481
25 caaattaaag aactccatgc cactcctcaa aaaaaaaaaa aaaaa 526

<210> 128

<211> 618

<212> DNA

30 <213> Homo sapiens

<220>

<221> CDS

<222> 179..427

35

<220>

<221> sig_peptide

<222> 179..298

<223> Von Heijne matrix

40 score 7.72883276007822
seq CLVVVTMATLSLA/RP

<400> 128
aagcgaagag atgggtctgc actttggagg agccggacac tgttgacttt cctgatgtga 60
45 aatctaccca ggaacaaaac accagtgact gcagcagcag cggcagcgcc tcggttcctg 120
agccaccgc aggtgaagg cattgcgcgt agtccatgcc cgtagaggaa gtgtgcag 178
atg gga tta acg tcc aca tgg aga tat gga aga gga ccg ggg att ggt 226
Met Gly Leu Thr Ser Thr Trp Arg Tyr Gly Arg Gly Pro Gly Ile Gly
-40 -35 -30 -25
50 acc gta acc atg gtc agc tgg ggt cgt ttc atc tgc ctg gtc gtg gtc 274
Thr Val Thr Met Val Ser Trp Gly Arg Phe Ile Cys Leu Val Val Val
-20 -15 -10
acc atg gca acc ttg tcc ctg gcc cgg ccc tcc ttc agt tta gtt gag 322
Thr Met Ala Thr Leu Ser Leu Ala Arg Pro Ser Phe Ser Leu Val Glu
-5 1 5
55 gat acc aca tta gag cca gaa gat gcc atc tca tcc gga gat gat gag 370
Asp Thr Thr Leu Glu Pro Glu Asp Ala Ile Ser Ser Gly Asp Asp Glu
10 15 20
gat gac acc gat ggt gcg gaa gat ttt gtc agt gag aac agt aac aac 418
60 Asp Asp Thr Asp Gly Ala Glu Asp Phe Val Ser Glu Asn Ser Asn Asn
25 30 35 40
aag agt aag taactgccc gctccgatgg tccccgagag aggagcatgg 467
Lys Ser Lys

agggaagttc	tgctgtcac	ctgtcttctt	gtcgactctt	ctgcgccatg	ctgtgtcccg	527
cggcccttgc	ctttcccccgc	tgtgtctact	ttcctgactt	tcaaacctga	gaataaacca	587
gtgttgctgc	acataaaaaa	aaaaaaaaaa	a			618

```
5 <210> 129
   <211> 776
   <212> DNA
   <213> Homo sapiens
```

```
10  <220>  
    <221> CDS  
    <222> 22..297
```

```

15 <220>
    <221> sig_peptide
    <222> 22..66
    <223> Von Heijne matrix
        score 4.68058603039206
        seq VLAGSLLGPTSRS/AA

```

20	<400> 129	
	actgcggggac ccactgcgga t atg gct gtc ttg gct gga tcc ctg ttg ggc	51
	Met Ala Val Leu Ala Gly Ser Leu Leu Gly	
	-15 -10	
25	ccc acg agt agg tcg gca gcg ttg ctg ggt ggc agg tgg ctc cag ccc	99
	Pro Thr Ser Arg Ser Ala Leu Leu Gly Gly Arg Trp Leu Gln Pro	
	-5 1 5 10	
	cgg gcc tgg ctg ggg ttc cca gac gcc tgg ggc ctc ccc acc ccg cag	147
	Arg Ala Trp Leu Gly Phe Pro Asp Ala Trp Gly Leu Pro Thr Pro Gln	
	15 20 25	
30	cag gcc cgg ggc aag gct cgc ggg aat gag tat cag ccg agc aat atc	195
	Gln Ala Arg Gly Lys Ala Arg Gly Asn Glu Tyr Gln Pro Ser Asn Ile	
	30 35 40	
	aaa cgc aag aac aag cac ggc tgg gtc cgg cgc ctg agc acg ccg gcc	243
35	Lys Arg Lys Asn Lys His Gly Trp Val Arg Arg Leu Ser Thr Pro Ala	
	45 50 55	
	ggc gtg cag gtc atc ctt cgc cga atg ctc aag ggc cgc aag tcg ctg	291
	Gly Val Gln Val Ile Leu Arg Arg Met Leu Lys Gly Arg Lys Ser Leu	
	60 65 70 75	
40	agc cat tgaggatcgc gacgcagtcg gcggggaccc tcatggaagc atcgccctcg	347
	Ser His	
	cctcggacct tgcctggcgc tattttttgca gggagctggg gaggcaggaac gcctcggacc	407
	tgagtgtctt ccatattgtg ggtttgaagt ctggatggga gccttgccaa gtcccttttt	467
	aggctttttta attaggaagc atttcgaacc tgcgcaacag accaaagaac agtacaaaga	527
45	acatccgtgt acccagtagc ctgactaccg actacctaca acccgccctt gccccatcct	587
	gagttctttt gaagctgata tcaggcatcg gattattttct tctgtaaata tttcagaatg	647
	tatctctcaata agatgagagc tcattaaaag ataattacaa agcttatcac atccaaaaga	707
	attatcacaata attttgaaat attattaaac gtgtaataaa tgttcaaagt tcaaaaaaaaa	767
	aaaaaaaaa	776

```
50      <210> 130
      <211> 998
      <212> DNA
      <213> Homo sapiens
```

55
 <220>
 <221> CDS
 <222> 9..845

```
60 <220>
    <221> sig_peptide
    <222> 9..134
    <223> Von Heijne matrix
```

score 6.13963522287438
seq RSLALAAAPSSNG/SP

```

<400> 130
5 aacgaaag atg gcg gcg ccc gta agg cgg acg ctg tta ggg gtg gcg ggg      50
   Met Ala Ala Pro Val Arg Arg Thr Leu Leu Gly Val Ala Gly
       -40               -35               -30

   ggt tgg cgg cgg ttc gag agg ctc tgg gcc ggc agt cta agc tct cgc      98
   Gly Trp Arg Arg Phe Glu Arg Leu Trp Ala Gly Ser Leu Ser Ser Arg
       -25               -20               -15

10 agc ctg gct ctt gca gcc gca ccc tca agc aac gga tcc cca tgg cgc      146
   Ser Leu Ala Leu Ala Ala Ala Pro Ser Ser Asn Gly Ser Pro Trp Arg
       -10               -5               1

   ttg ttg ggc gcg ttg tgc ctg cag cgg cca cct gta gtc tcc aag ccg      194
   Leu Leu Gly Ala Leu Cys Leu Gln Arg Pro Pro Val Val Ser Lys Pro
   5               10               15               20

   ttg acc cca ttg cag gaa gag atg gcg tct cta ctg cag cag att gag      242
   Leu Thr Pro Leu Gln Glu Glu Met Ala Ser Leu Leu Gln Gln Ile Glu
       25               30               35

20 ata gag aga agc ctg tat tca gac cac gag ctt cgt gct ctg gat gaa      290
   Ile Glu Arg Ser Leu Tyr Ser Asp His Glu Leu Arg Ala Leu Asp Glu
       40               45               50

   aac cag cga ctg gca aag aag aaa gct gac ctt cat gat gaa gaa gat      338
   Asn Gln Arg Leu Ala Lys Lys Lys Ala Asp Leu His Asp Glu Glu Asp
       55               60               65

25 gaa cag gat ata ttg ctg gcg caa gat ttg gaa gat atg tgg gag cag      386
   Glu Gln Asp Ile Leu Leu Ala Gln Asp Leu Glu Asp Met Trp Glu Gln
       70               75               80

   aaa ttt cta cag ttc aaa ctt gga gct cgc ata aca gaa gct gat gaa      434
   Lys Phe Leu Gln Phe Lys Leu Gly Ala Arg Ile Thr Glu Ala Asp Glu
   85               90               95               100

   aag aat gac cga aca tcc ctg aac agg aac cta gac agg aac ctt gtc      482
   Lys Asn Asp Arg Thr Ser Leu Asn Arg Asn Leu Asp Arg Asn Leu Val
       105               110               115

35 ctg tta gtc aga gag aag ttt gga gac cag gat gtt tgg ata ctg ccc      530
   Leu Leu Val Arg Glu Lys Phe Gly Asp Gln Asp Val Trp Ile Leu Pro
       120               125               130

   cag gca gag tgg cag cct ggg gag acc ctt cga gga aca gct gaa cga      578
   Gln Ala Glu Trp Gln Pro Gly Glu Thr Leu Arg Gly Thr Ala Glu Arg
       135               140               145

40 acc ctg gcc aca ctc tca gaa aac aac atg gaa gcc aag ttc cta gga      626
   Thr Leu Ala Thr Leu Ser Glu Asn Asn Met Glu Ala Lys Phe Leu Gly
       150               155               160

   aat gca ccc tgt ggg cac tac aca ttc aag ttc ccc cag gca atg cgg      674
   Asn Ala Pro Cys Gly His Tyr Thr Phe Lys Phe Pro Gln Ala Met Arg
   165               170               175               180

   aca gag agt aac ctc gga gcc aag gtg ttc ttc ttc aaa gca ctg cta      722
   Thr Glu Ser Asn Leu Gly Ala Lys Val Phe Phe Phe Lys Ala Leu Leu
       185               190               195

50 tta act gga gac ttt tcc cag gct ggg aat aag ggc cat cat gtg tgg      770
   Leu Thr Gly Asp Phe Ser Gln Ala Gly Asn Lys Gly His His Val Trp
       200               205               210

   gtc att aag gat gag ctg ggt gac tat ttg aaa cca aaa tac ctg gcc      818
   Val Ile Lys Asp Glu Leu Gly Asp Tyr Leu Lys Pro Lys Tyr Leu Ala
       215               220               225

55 caa gtt agg agg ttt gtt tca gac ctc tgatgggccc agctgcctgt      865
   Gln Val Arg Arg Phe Val Ser Asp Leu
       230               235

   ggacggtgct cagacaagtc tgggattaga gcctcaagga cattgtgtga ttgcctcaca      925
60 tttgcaggta atatcaagca gcaaactaaa ttctgagaaa taaacgagtc tattacaaa      985
   aaaaaaaaaa aaa
       230               235

```

<210> 131

<211> 779
 <212> DNA
 <213> Homo sapiens

5 <220>
 <221> CDS
 <222> 27..578

<220>
 10 <221> sig_peptide
 <222> 27..119
 <223> Von Heijne matrix
 score 4.50637135496675
 seq TALMVGAASLLEG/RP

15 <400> 131
 atcttttctg actggccctg cagagg atg gca tgc acc act act gcc ccc gcc 53
 Met Ala Cys Thr Thr Thr Ala Pro Ala
 -30 -25
 20 cag gaa cac atg ctt ctc acc cct ctc act gct ctg atg gtg ggg gct 101
 Gln Glu His Met Leu Leu Thr Pro Leu Thr Ala Leu Met Val Gly Ala
 -20 -15 -10
 gct tct ctg ctt gag ggc cgg cca cag atc tca gct cca tac tcc cga 149
 Ala Ser Leu Leu Glu Gly Arg Pro Gln Ile Ser Ala Pro Tyr Ser Arg
 25 -5 1 5 10
 gct gca tgt tgc agc cct ggg gca ctg gga tgt cct gca gct cgg gtt 197
 Ala Ala Cys Cys Ser Pro Gly Ala Leu Gly Cys Pro Ala Ala Arg Val
 15 20 25
 ggg att ctg gat ctg atg tat tcc tgg gtt gcc agg aaa gtg ctc agg 245
 30 Gly Ile Leu Asp Leu Met Tyr Ser Trp Val Ala Arg Lys Val Leu Arg
 30 35 40
 tgc agc aat act ggg ctg cag ggg ctg cac tgt gca cca gct tat gca 293
 Cys Ser Asn Thr Gly Leu Gln Gly Leu His Cys Ala Pro Ala Tyr Ala
 45 50 55
 35 gca cag ctt ggt atg gac cct ggg agg ggc caa cga gca gga ggg cct 341
 Ala Gln Leu Gly Met Asp Pro Gly Arg Gly Gln Arg Ala Gly Gly Pro
 60 65 70
 gta gag cag aca tac ttc agt ccc atg ggg aag ctg ccc act ctt tcg 389
 Val Glu Gln Thr Tyr Phe Ser Pro Met Gly Lys Leu Pro Thr Leu Ser
 40 75 80 85 90
 tgg ctg gaa ggc tgt aca gca gtc atg acg ctg gca tct gct tgg ctt 437
 Trp Leu Glu Gly Cys Thr Ala Val Met Thr Leu Ala Ser Ala Trp Leu
 95 100 105
 ctg ggg agc cct cgg gaa act tac aat cat gag aag gtg aag gag aag 485
 45 Leu Gly Ser Pro Arg Glu Thr Tyr Asn His Glu Lys Val Lys Glu Lys
 110 115 120
 cag tgt cca ttc tcc agt atg gtt ttg ggg gag tat ggc ttc cta cct 533
 Gln Cys Pro Phe Ser Ser Met Val Leu Gly Glu Tyr Gly Phe Leu Pro
 125 130 135
 50 act gtg gac cac ctg tca act ctg ggc tgt aac atg aga gaa ttg 578
 Thr Val Asp His Leu Ser Thr Leu Gly Cys Asn Met Arg Glu Leu
 140 145 150
 tgaacttctg tcttggttga gccatggttt cattctcttt ttcagccatg tagcctgtgc 638
 tgtaactcag taccacatta gcaactagt aaagtcaatg tgggtaaatt tgtcattctt 698
 55 caggttagaa catttcttcc ttttattctt gtgttttttg ctaaataaac tgggaaatta 758
 tagtaaaaaa aaaaaaaaaa a 779

<210> 132
 <211> 1025
 60 <212> DNA
 <213> Homo sapiens
 <220>

```

<221> CDS
<222> 408..710

<220>
5 <221> sig_peptide
  <222> 408..533
  <223> Von Heijne matrix
        score 5.66440183652506
        seq QLCFHLSWLYSWA/SQ

10
  <400> 132
  atggtttgtt gtagattcca tgcctcttgg gatcagtcac tgtggccatg catgtttggc      60
  cacatgatta atccagtctg ggtcatgacc ttttcttcat ccaaaacaag gtgatgggaa      120
  gacaaaaaca atagctacta caaacaatag gagtttataa ttatgtgctg atgtattcga      180
15 agatgtgttg acagtcgtga gtgtgtatcc taggaaaggc gagctggact ctgtctccat      240
  ggtggctctc accccaggga cctaggaaca gcctgtcacc acacaattac ttttataacc      300
  ctggagatga aaatctcctt gtcctcaaaa tacttccaga agaacaacca gatgggaagg      360
  accttggttg ggactctttc cagttcactt ggggcagagg gaatttta atg gct cac      416
                                     Met Ala His
                                     -40
20 gta gct gaa aag gat ggg cta gat tgg gct tca ggc tgc atc cca gga      464
  Val Ala Glu Lys Asp Gly Leu Asp Trp Ala Ser Gly Cys Ile Pro Gly
                                     -35 -30 -25
  ctc caa aca ggg atc tgt ctc ttt ggc tct cag ctc tgc ttt cat ttg      512
25 Leu Gln Thr Gly Ile Cys Leu Phe Gly Ser Gln Leu Cys Phe His Leu
                                     -20 -15 -10
  agt tgg ctt tat tct tgg gct tca cag tgt ggc ccc aca gca cca gtt      560
  Ser Trp Leu Tyr Ser Trp Ala Ser Gln Cys Gly Pro Thr Ala Pro Val
                                     -5 1 5
30 att gat aaa aag agc tcc cct ttg ctg aca gaa ctg ctg gat ttg gtt      608
  Ile Asp Lys Lys Ser Ser Pro Leu Leu Thr Glu Leu Leu Asp Leu Val
  10 15 20 25
  ctc att ggt cca gac gag gaa ggt atc cag cct caa gtc atc att gtg      656
  Leu Ile Gly Pro Asp Glu Glu Gly Ile Gln Pro Gln Val Ile Ile Val
35 30 35 40
  gcc agg aag atg gaa tac acc aaa tgg aca ggc ctg gca tgt acc cac      704
  Ala Arg Lys Met Glu Tyr Thr Lys Trp Thr Gly Leu Ala Cys Thr His
  45 50 55
  aga gac tgagagttgg tgctggtggt tgtggtggca gatgatatta cctgaagaag      760
40 Arg Asp
  ggacgaatgg gtgctgggca ggacaaagca tcagctgtcc agttcaggcc tctcctcttt      820
  ccctggtgtc ttcattttcc tccgtctccc tgctgtccct taccctctgc ccaatctcat      880
  tactcctggt cttgggaggt gccttctgag gatactccac tgggggtacc tgagcctgga      940
  ttagagggca gggggaggat attgcctagc caaagtgggt gttcaataaa gaaccatttg      1000
45 gagatggcaa aaaaaaaaaa aaaaa
                                     1025

<210> 133
<211> 607
<212> DNA
50 <213> Homo sapiens

<220>
<221> CDS
<222> 247..501

55
<220>
<221> sig_peptide
<222> 247..306
<223> Von Heijne matrix
60 score 6.43040298500966
  seq LLLVTLVASTVPG/NS

<400> 133

```

```

tgttacaaat attccctatg atctctcctt taaatattct tatcaggata ttggaaattc      60
ttgattttca caactctgct tcagtgccat atgttttagct ttttgtcttc tgaattaatt      120
gggcttctga tgggtccctag aggtatcagc tactcagtc gaaaacatac atggggaaga      180
aactgaagtt catgccacaa actgtagcag ctttgggaaca gaagggacca gacaacctca      240
5  aggaga atg ggc cca aat acc aaa aat tta ctc ttg gtg acc ctt gtt      288
    Met Gly Pro Asn Thr Lys Asn Leu Leu Leu Val Thr Leu Val
      -20                -15                -10

gct tct act gta cca ggc aac tct ctt ggg cag gat ttt act ttt gca      336
Ala Ser Thr Val Pro Gly Asn Ser Leu Gly Gln Asp Phe Thr Phe Ala
10  -5                1                5                10

cac tta gaa aga tcc tgc acc agg gaa aat cgg tct cct ggg gag gta      384
His Leu Glu Arg Ser Cys Thr Arg Glu Asn Arg Ser Pro Gly Glu Val
      15                20                25

ttc cag caa cca tgc aag tct gga ggc ggg ggg gtt gga gaa cca aat      432
15  Phe Gln Gln Pro Cys Lys Ser Gly Gly Gly Gly Val Gly Glu Pro Asn
      30                35                40

gcc caa ggg cag cta ctt agc cag cac cca cta cct gcc ttc att aat      480
Ala Gln Gly Gln Leu Leu Ser Gln His Pro Leu Pro Ala Phe Ile Asn
      45                50                55

20  tgt tct cac ggg cag gcc ttt tgaaccaccc tggtacagaa caccaaccct      531
    Cys Ser His Gly Gln Ala Phe
      60                65

ggtgcttttag gctgtctgtg ccatttctag gcaatgaacg agtagttact gtaccaaccc      591
aaaaa         aaaaaa      607

25  <210> 134
    <211> 774
    <212> DNA
    <213> Homo sapiens

30  <220>
    <221> CDS
    <222> 333..602

35  <220>
    <221> sig_peptide
    <222> 333..416
    <223> Von Heijne matrix
          score 4.79986448293481

40  seq VPALPLLSSLCMA/MV

    <400> 134
ctcttcagtc cggggccttgg ttgaacggac tcaccaggaa acgtgacttt cgtgtccgac      60
ctctgctgta tcaggattcg attccttggtg ttaaacaaga caacgctgaa ggctcgggtgc      120
45  agcagccctg caaaggtttt tccagcgctc ttgggagggtg ggctgtgccc tgcctggccc
    acctggccca cctggcccac cattacctga agggaagcat gaacagcctt tgacgtggga      240
gtggcgactg ctgagaggga actgtctgta cacaagcaat gtagccttat gggacctgag      300
tgagagcccca acccacgcag ggcgtgktct tc atg gct ttt cct ggc caa tct      353
          Met Ala Phe Pro Gly Gln Ser
          -25

50  gat acc aag atg cag tgg cca gaa gta cct gca ctt cca ctc ctg tca      401
    Asp Thr Lys Met Gln Trp Pro Glu Val Pro Ala Leu Pro Leu Leu Ser
      -20                -15                -10

agt ctc tgc atg gct atg gtg agg aag agc tct gca ctg ggc aag gaa      449
55  Ser Leu Cys Met Ala Met Val Arg Lys Ser Ser Ala Leu Gly Lys Glu
      -5                1                5                10

gtt ggc cgt cga gtg aag gaa atg gtg atg ctg gtg gcc cct ttc cgg      497
Val Gly Arg Arg Val Lys Glu Met Val Met Leu Val Ala Pro Phe Arg
      15                20                25

60  cag tca agt tcc cta tca agg aca ttc agt tct cgg aaa gtg gtg aag      545
    Gln Ser Ser Ser Leu Ser Arg Thr Phe Ser Ser Arg Lys Val Val Lys
      30                35                40

gca cat gct tcc ctg cat ggt gcc cgc ctc tct cca ctc tct aga aat      593

```

Ala His Ala Ser Leu His Gly Ala Arg Leu Ser Pro Leu Ser Arg Asn
 45 50 55
 att aga ggc taggctgctg ctgtatgtca gggctagtcc ctcttctatg 642
 Ile Arg Gly
 5 60
 aatccagaat aactctgaag aagccgagta acaggcatga agtgaagaga aatcgctgta 702
 acaggaagac agcaaagcag atgctaataa ccacactatt taacgaactg gaaccaacaa 762
 aaaaaaaaaa aa 774

10 <210> 135
 <211> 611
 <212> DNA
 <213> Homo sapiens

15 <220>
 <221> CDS
 <222> 110..376

20 <220>
 <221> sig_peptide
 <222> 110..208
 <223> Von Heijne matrix
 score 3.64796206065748
 seq LVPHSPLPGALSS/AP

25 <220>
 <221> misc_feature
 <222> 347
 <223> n=a, g, c or t

30 <400> 135
 tcttgtcaac actgcccact cagcgaggaa gcagccgcga cgcccacact tcttgttgga 60
 gcctgcgcag agccagaggc ctccagaagcc acaggaacat ggcctaggc atg gct cag 118
 Met Ala Gln

35 cca gca gcc ccc tcc ctg acg cgg ccc ttc ctg gca gag gcc ccg aca 166
 Pro Ala Ala Pro Ser Leu Thr Arg Pro Phe Leu Ala Glu Ala Pro Thr
 -30 -25 -20 -15
 gca ctg gtc cca cac agc ccc ctg cct ggg gcc ctg tca agc gcc cct 214
 Ala Leu Val Pro His Ser Pro Leu Pro Gly Ala Leu Ser Ser Ala Pro
 -10 -5 1

40 ggc ccg aag cag ccc ccg acg gca agc aca ggc ccg gag ctg ctg ctg 262
 Gly Pro Lys Gln Pro Pro Thr Ala Ser Thr Gly Pro Glu Leu Leu Leu
 5 10 15
 ctg cct ctt tcc tcc ttc atg ccc tgc ggg gcg gct gca cca gcc agg 310
 45 Leu Pro Leu Ser Ser Phe Met Pro Cys Gly Ala Ala Pro Ala Arg
 20 25 30
 gtg tca tca cag cgg gct act cct agg gat aag ccc ncc ggt ccc ctc 358
 Val Ser Ser Gln Arg Ala Thr Pro Arg Asp Lys Pro Xaa Gly Pro Leu
 35 40 45 50

50 atc cct ggc cag tgt ccc tgacccccat ctactccttc ctggggactt 406
 Ile Pro Gly Gln Cys Pro
 55
 ctcagcgcca gccattggc gcctgcgttg cccgcattcca ggccctgcgg caggccctgt 466
 gctagcgtgt tcgcaccagg aacgcaggtg ctgggctgtc ggggaggcct caggccacct 526
 55 ccaggaacag aacacagttt taagtttgat tttttttatt tcaaaatgct ttgcaattaa 586
 atgaattact gttcaaaaaa aaaaa 611

<210> 136
 <211> 925
 60 <212> DNA
 <213> Homo sapiens

<220>

<221> CDS
<222> 22..417

<220>

5 <221> sig_peptide
<222> 22..66
<223> Von Heijne matrix
score 5.47092708754574
seq RVLCPAAGAVRA/LR

10

<400> 136
agtcgaggag tcaaggcagc a atg aat cgt gtc ttg tgt gcc ccg gcg gcc 51
Met Asn Arg Val Leu Cys Ala Pro Ala Ala
-15 -10
15 ggg gcc gtc cgg gcg ctg agg ctc ata ggc tgg gct tcc cga agc ctt 99
Gly Ala Val Arg Ala Leu Arg Leu Ile Gly Trp Ala Ser Arg Ser Leu
-5 1 5 10
cat ccg ttg ccc ggt tcc cgg gat cgg gcc cac cct gcc gcc gag gaa 147
His Pro Leu Pro Gly Ser Arg Asp Arg Ala His Pro Ala Ala Glu Glu
20 15 20 25
gag gac gac cct gac cgc ccc att gag ttt tcc tcc agc aaa gcc aac 195
Glu Asp Asp Pro Asp Arg Pro Ile Glu Phe Ser Ser Ser Lys Ala Asn
30 35 40
cct cac cgc tgg tcg gtg ggc cat acc atg gga aag gga cat cag cgg 243
25 Pro His Arg Trp Ser Val Gly His Thr Met Gly Lys Gly His Gln Arg
45 50 55
ccc tgg tgg aag gtg ctg ccc ctc agc tgc ttc ctc gtg gcg ctg atc 291
Pro Trp Trp Lys Val Leu Pro Leu Ser Cys Phe Leu Val Ala Leu Ile
60 65 70 75
30 atc tgg tgc tac ctg agg gag gag agc gag gcg gac cag tgg ttg aga 339
Ile Trp Cys Tyr Leu Arg Glu Glu Ser Glu Ala Asp Gln Trp Leu Arg
80 85 90
cag gtg tgg gga gag gtg cca gag ccc agt gat cgt tct gag gag cct 387
Gln Val Trp Gly Glu Val Pro Glu Pro Ser Asp Arg Ser Glu Glu Pro
35 95 100 105
gag act cca gct gcc tac aga gcg aga act tgacggggtg cccgctgggg 437
Glu Thr Pro Ala Ala Tyr Arg Ala Arg Thr
110 115
ctggcaggaa gggagccgac agccgccctt cggatttgat gtcacgtttg cccgtgactg 497
40 tcctggctat gcgtgcgtcc tcagcactga aggacttggc tgggtggatgg ggcacttggc 557
tatgctgatt cgcgtgaagg cggagcagaa tctcagcaga tcggaaactg ctctcgcct 617
ggctcttgat gtccaaggat tccatcggca agacttctca gatccttggg gaaggtttca 677
gttgcaactgt atgctgttgg atttgccaaag tctttgtata acataatcat gtttccaaag 737
cacttctggt gacacttgtc atccagtgtt agtttgcagg taatttgctt tctgagatag 797
45 aatatctggc agaagtgtga aactgtattg catgctgcgg cctgtgcaag gaacacttcc 857
acatgtgagt ttacacaac aacaaatgaa aataaatttt aattttataa taaaaaaaaa 917
aaaaaaaaa 925

<210> 137

50 <211> 674

<212> DNA

<213> Homo sapiens

<220>

55 <221> CDS
<222> 62..367

<220>

60 <221> sig_peptide
<222> 62..103
<223> Von Heijne matrix
score 4.49063834776683
seq FLTALLWRGRIPG/RQ

<400> 137
acgccacggc gtctgctggc ggccgcggag acgcagagtc ttgagcagcg cggcaggcac 60
c atg ttc ctg act gcg ctc ctc tgg cgc ggc cgc att ccc ggc cgt cag 109
5 Met Phe Leu Thr Ala Leu Leu Trp Arg Gly Arg Ile Pro Gly Arg Gln
-10 -5 1
tgg atc ggg aag cac cgg cgg ccg cgg ttc gtg tcg ttg cgc gcc aag 157
Trp Ile Gly Lys His Arg Arg Pro Arg Phe Val Ser Leu Arg Ala Lys
5 10 15
10 cag aac atg atc cgc cgc ctg gag atc gat gcg gag aac cat tac tgg 205
Gln Asn Met Ile Arg Arg Leu Glu Ile Asp Ala Glu Asn His Tyr Trp
20 25 30
ctg agc atg ccc tac atg acc cgg gag cag gag cgc ggc cac gcc gsg 253
Leu Ser Met Pro Tyr Met Thr Arg Glu Gln Glu Arg Gly His Ala Xaa
15 35 40 45 50
dtg cgc agg agg gag gcc ttc gag gcc ata aag gcg gcc gcc act tcc 301
Xaa Arg Arg Arg Glu Ala Phe Glu Ala Ile Lys Ala Ala Ala Thr Ser
55 60 65
aag ttc ccc ccg cat aga ttc att gcg gac cag ctc gac cat ctc aat 349
20 Lys Phe Pro Pro His Arg Phe Ile Ala Asp Gln Leu Asp His Leu Asn
70 75 80
gtc acc aag aaa tgg tcc taatcctgag tagtcaccct tggattttat 397
Val Thr Lys Lys Trp Ser
85
25 ggatcacgga gctgaccatc tttacctggt cctggaactg aaaaactgta gcttgtgtga 457
aaatgagcct ttggaccagt ctttattaaa acaaacaaac atgagtagtc tgcataatcga 517
atatctagag ctctaaaccc cccaatactt aaaagtctaa ttgctgtcct gtggtttcat 577
tagtctgata ggaagatagg gatttcctca gtcacagatg atattttgaa ggaaagctgc 637
aataaagcca caatgattcg aaaaaaaaaa aaaaaaa 674
30
<210> 138
<211> 1725
<212> DNA
<213> Homo sapiens
35
<220>
<221> CDS
<222> 107..1618
40
<220>
<221> sig_peptide
<222> 107..178
<223> Von Heijne matrix
score 6.19650168602189
45 seq LGLYSLVLSGALA/YA
<400> 138
agagctcagc cggtcgcacg gacggacagt tggaagccgg accccagagc ctgaggtggg 60
cagtgtgccca gggctcccttg cggcctcctc aagccctgtc caggct atg ggc atc 115
50 Met Gly Ile
aag aca gca ttg ccg gcg gct gag ctg ggc ctc tac tct ctg gtg ctg 163
Lys Thr Ala Leu Pro Ala Ala Glu Leu Gly Leu Tyr Ser Leu Val Leu
-20 -15 -10
agt ggg gcc ctg gcc tat gct ggc cgg ggc ctc ctt gag gct tca caa 211
55 Ser Gly Ala Leu Ala Tyr Ala Gly Arg Gly Leu Leu Glu Ala Ser Gln
-5 1 5 10
gat ggg gcc cac agg aag gcc ttc cgg gag tct gtg cga cct ggc tgg 259
Asp Gly Ala His Arg Lys Ala Phe Arg Glu Ser Val Arg Pro Gly Trp
15 20 25
60 gag tac att ggc cgg aag atg gat gtg gct gac ttc gag tgg gtg atg 307
Glu Tyr Ile Gly Arg Lys Met Asp Val Ala Asp Phe Glu Trp Val Met
30 35 40
tgg ttc acc tcc ttt cgc aac gtc atc atc ttt gcc ctc tcc gga cat 355

	Trp	Phe	Thr	Ser	Phe	Arg	Asn	Val	Ile	Ile	Phe	Ala	Leu	Ser	Gly	His	
	45						50				55						
	gtg	ctg	ttt	gct	aaa	ctc	tgc	acg	atg	gtt	gcc	cca	aag	ctc	cgc	tcc	403
	Val	Leu	Phe	Ala	Lys	Leu	Cys	Thr	Met	Val	Ala	Pro	Lys	Leu	Arg	Ser	
5	60				65					70					75		
	tgg	atg	tat	gct	gtg	tac	ggg	gcc	ttg	gct	gtg	atg	ggc	aca	atg	ggc	451
	Trp	Met	Tyr	Ala	Val	Tyr	Gly	Ala	Leu	Ala	Val	Met	Gly	Thr	Met	Gly	
				80				85						90			
	cct	tgg	tac	ctg	ctg	ctg	ctg	ctt	ggg	cac	tgt	gtg	ggc	ctc	tat	gtg	499
10	Pro	Trp	Tyr	Leu	Leu	Leu	Leu	Leu	Gly	His	Cys	Val	Gly	Leu	Tyr	Val	
				95				100						105			
	gcc	tgc	ctt	ttg	ggc	cag	ccc	tgg	ctc	tgt	ctt	ggc	ctt	ggc	ttg	gcc	547
	Ala	Ser	Leu	Leu	Gly	Gln	Pro	Trp	Leu	Cys	Leu	Gly	Leu	Gly	Leu	Ala	
			110				115					120					
15	agc	ctg	gcc	tcc	ttc	aag	atg	gac	ccc	cta	atc	tct	tgg	cag	agc	ggg	595
	Ser	Leu	Ala	Ser	Phe	Lys	Met	Asp	Pro	Leu	Ile	Ser	Trp	Gln	Ser	Gly	
			125				130					135					
	ttt	gta	aca	ggc	act	ttt	gat	ctt	caa	gag	gtg	ctg	ttt	cat	ggg	ggc	643
	Phe	Val	Thr	Gly	Thr	Phe	Asp	Leu	Gln	Glu	Val	Leu	Phe	His	Gly	Gly	
20	140					145					150				155		
	agc	agc	ttc	aca	gtg	ctg	cgt	tgc	acc	agc	ttt	gca	ctg	gag	agc	tgt	691
	Ser	Ser	Phe	Thr	Val	Leu	Arg	Cys	Thr	Ser	Phe	Ala	Leu	Glu	Ser	Cys	
				160				165						170			
	gcc	cac	cct	gac	cgc	cac	tac	tcc	tta	gct	gac	ctg	ctc	aag	tac	agc	739
25	Ala	His	Pro	Asp	Arg	His	Tyr	Ser	Leu	Ala	Asp	Leu	Leu	Lys	Tyr	Ser	
				175				180						185			
	ttc	tac	ctg	ccc	ttc	ttc	ttc	ttc	ggg	ccc	atc	atg	acc	ttt	gat	cgc	787
	Phe	Tyr	Leu	Pro	Phe	Phe	Phe	Phe	Gly	Pro	Ile	Met	Thr	Phe	Asp	Arg	
			190					195					200				
30	ttc	cat	gct	cag	gtg	agc	cag	gtg	gag	cca	gtg	aga	cgc	gag	ggg	gag	835
	Phe	His	Ala	Gln	Val	Ser	Gln	Val	Glu	Pro	Val	Arg	Arg	Glu	Gly	Glu	
			205				210					215					
	ctg	tgg	cac	atc	cga	gcc	cag	gca	ggc	cta	agc	gtg	gtg	gcc	atc	atg	883
	Leu	Trp	His	Ile	Arg	Ala	Gln	Ala	Gly	Leu	Ser	Val	Val	Ala	Ile	Met	
35	220					225					230				235		
	gcc	gtc	gac	atc	ttc	ttt	cac	ttc	ttc	tac	atc	ctc	act	atc	ccc	agc	931
	Ala	Val	Asp	Ile	Phe	His	Phe	Phe	Phe	Tyr	Ile	Leu	Thr	Ile	Pro	Ser	
				240				245						250			
	gac	ctc	aag	ttc	gcc	aac	cgc	ctc	cca	gac	att	gcc	ctc	gct	ggc	cta	979
40	Asp	Leu	Lys	Phe	Ala	Asn	Arg	Leu	Pro	Asp	Ile	Ala	Leu	Ala	Gly	Leu	
				255				260						265			
	gcc	tat	tca	aac	ctg	gtg	tat	gac	tgg	gtg	aag	gcg	gcc	gtc	ctc	ttt	1027
	Ala	Tyr	Ser	Asn	Leu	Val	Tyr	Asp	Trp	Val	Lys	Ala	Ala	Val	Leu	Phe	
			270					275					280				
45	ggg	gtt	gtc	aac	act	gtg	gca	tgc	ctc	gac	cac	ctg	gac	cca	ccc	cag	1075
	Gly	Val	Val	Asn	Thr	Val	Ala	Cys	Leu	Asp	His	Leu	Asp	Pro	Pro	Gln	
			285				290					295					
	cct	ccc	aag	tgc	atc	acc	gca	ctc	tac	gtc	ttt	gcg	gaa	acg	cac	ttt	1123
	Pro	Pro	Lys	Cys	Ile	Thr	Ala	Leu	Tyr	Val	Phe	Ala	Glu	Thr	His	Phe	
50	300					305					310				315		
	gac	cgt	ggc	atc	aac	gac	tgg	ctt	tgc	aaa	tat	gtg	tat	aac	cac	att	1171
	Asp	Arg	Gly	Ile	Asn	Asp	Trp	Leu	Cys	Lys	Tyr	Val	Tyr	Asn	His	Ile	
				320				325						330			
	ggg	ggg	gag	cat	tcc	gct	gtg	atc	cca	gag	ctg	gca	gcc	aca	gtg	gcc	1219
55	Gly	Gly	Glu	His	Ser	Ala	Val	Ile	Pro	Glu	Leu	Ala	Ala	Thr	Val	Ala	
				335				340						345			
	aca	ttt	gcc	atc	acc	aca	ctg	tgg	ctt	ggg	cct	tgt	gac	att	gtc	tac	1267
	Thr	Phe	Ala	Ile	Thr	Thr	Leu	Trp	Leu	Gly	Pro	Cys	Asp	Ile	Val	Tyr	
			350				355						360				
60	ctg	tgg	tca	ttc	ctt	aac	tgc	ttt	ggc	ctc	aac	ttt	gag	ctc	tgg	atg	1315
	Leu	Trp	Ser	Phe	Leu	Asn	Cys	Phe	Gly	Leu	Asn	Phe	Glu	Leu	Trp	Met	
			365				370					375					
	caa	aaa	ctg	gca	gag	tgg	ggg	ccc	cta	gca	cga	att	gag	gcc	tct	ctg	1363

```

Gln Lys Leu Ala Glu Trp Gly Pro Leu Ala Arg Ile Glu Ala Ser Leu
380          385          390          395
tca gtg cag atg tcc cgt agg gtc cgg gcc ctg ttt gga gcc atg aac      1411
Ser Val Gln Met Ser Arg Arg Val Arg Ala Leu Phe Gly Ala Met Asn
5          400          405          410
ttc tgg gcc atc atc atg tac aac ctt gtg agc ctg aac agc ctc aaa      1459
Phe Trp Ala Ile Ile Met Tyr Asn Leu Val Ser Leu Asn Ser Leu Lys
          415          420          425
ttc aca gag ctg gtt gcc cgg cgc ctg cta ctc aca ggg ttc ccc cag      1507
10 Phe Thr Glu Leu Val Ala Arg Arg Leu Leu Leu Thr Gly Phe Pro Gln
          430          435          440
acc acg ctg tcc atc ctg ttt gtc acc tac tgt ggc gtc cag ctg gta      1555
Thr Thr Leu Ser Ile Leu Phe Val Thr Tyr Cys Gly Val Gln Leu Val
          445          450          455
15 aag gag cgt gag cga acc ttg gca ctg gag gag gag gag cag aag cag gac      1603
Lys Glu Arg Glu Arg Thr Leu Ala Leu Glu Glu Glu Gln Lys Gln Asp
460          465          470          475
aaa gag aag ccg gag taggaggggag cgggtagagg gatgggctct gctcagctat      1658
Lys Glu Lys Pro Glu
20          480
tcttggggcca gatggggcct gaccgataga ataaaagact tttctacaac aaaaaaaaaa      1718
aaaaaaaaa      1725

<210> 139
25 <211> 1474
    <212> DNA
    <213> Homo sapiens

    <220>
30 <221> CDS
    <222> 16..471

    <220>
    <221> sig_peptide
35 <222> 16..93
    <223> Von Heijne matrix
        score 5.809301698725
        seq FCVCVIAIGVVQA/LI

40 <400> 139
    tacacgtttt cgтта atg gtg acc ttc cct gat gtg cct ctg ggc atc ttc      51
        Met Val Thr Phe Pro Asp Val Pro Leu Gly Ile Phe
        -25          -20          -15
45 ttg ttc tgt gtg tgt gtg atc gcc atc ggg gtc gtg cag gca ctg att      99
    Leu Phe Cys Val Cys Val Ile Ala Ile Gly Val Val Gln Ala Leu Ile
        -10          -5          1
    gta ggg tac gca ttc cac ttc ccg cac ctg ctg agc ccg cag atc cag      147
    Val Gly Tyr Ala Phe His Phe Pro His Leu Leu Ser Pro Gln Ile Gln
        5          10          15
50 cgc tct gcc cac agg gct ctg tac cga cga cac gtc ctg ggc atc gtc      195
    Arg Ser Ala His Arg Ala Leu Tyr Arg Arg His Val Leu Gly Ile Val
        20          25          30
    ctc caa ggc ccg gcc ctg tgc ttt gca gcg gcc atc ttc tct ctc ttc      243
    Leu Gln Gly Pro Ala Leu Cys Phe Ala Ala Ile Phe Ser Leu Phe
55 35          40          45          50
    ttt gtc ccc ttg tct tac ctg ctg atg gtg act gtc atc ctc ctc ccc      291
    Phe Val Pro Leu Ser Tyr Leu Leu Met Val Thr Val Ile Leu Leu Pro
        55          60          65
60 tat gtc agc aag gtc acc ggc tgg tgc aga gac agg ctc ctg ggc cac      339
    Tyr Val Ser Lys Val Thr Gly Trp Cys Arg Asp Arg Leu Leu Gly His
        70          75          80
    agg gag ccc tcg gct cac cca gtg gaa gtc ttc tcg ttt gac ctc cac      387
    Arg Glu Pro Ser Ala His Pro Val Glu Val Phe Ser Phe Asp Leu His

```

```

      85              90              95
gag cca ctc agc aag gag cgc gtg gaa gcc ttc agc gac gga gtc tac      435
Glu Pro Leu Ser Lys Glu Arg Val Glu Ala Phe Ser Asp Gly Val Tyr
      100              105              110
5  gcc atc gtg gcc acg ctt ctc atc ctg gac atc tgg tgaggacccc      481
Ala Ile Val Ala Thr Leu Leu Ile Leu Asp Ile Trp
      115              120              125
gcgtcacctg cccagctat caggtggcca atgtgtcttg agtccctggc gtctcatcct      541
ggaaacccca gaaaggcaca ggggtcttgg ctccaccctc ctctggatgc ctagagtttt      601
10  gtgtgaggtc agggcagccc ccacttcagg gaggacaacc ttcccggcgg cccctccctt      661
cccagcggcc cctcccttcc cagaggctcc caccccaagc acagccgagg atgggggtgcc      721
agggtagggt cagcaccagc agccaactgc tctcctcact cctctcagag gggtcagca      781
gccatgggta tccccctgcc ccaggcctca cccctgcccc aacaccagcc cctcctagtc      841
cctagtcctt cccattccct ccggctccct cccagtgtccc cccatcgctt cgcagccctt      901
15  tctgtccctt ttggtgggt gttgttctt tccagcgtct gctcctccgc ggctcatct      961
gcctcttcgt ctgttagagc gcgcgtctcg tctcagtcgt caggtttttg gtttttgttg      1021
ggtttttttt tttttttttt tttgagacag tctgtgtgtg tcgcccaggc tggagtatag      1081
tggtcaagc tcagctcact gcaacctccg cctcccagggt tcaagcaatt ctctgcctc      1141
agcctcccaa gtagtggga ttacaagcac ccaccaccat gcccagctaa ctttttgcct      1201
20  ttttaataga gatgaggtt caccaagttg gccagggttg tcttgaactc ctgacctcag      1261
gtgatctgcc cacctcggcc tcccaaagtg ctgggattac aggtgtaagc caccgtgccc      1321
ggccatcgta atgtttgaat ttgctttttt acatcttcca tccttttgga gtgtcttggt      1381
ccctcgtcat agttcagcac tgtgaccacc ttgggggttag acactatggt tttatatcct      1441
gtacttgata ttctccaaaa aaaaaaaaaa aaa      1474
25
<210> 140
<211> 653
<212> DNA
<213> Homo sapiens
30
<220>
<221> CDS
<222> 222..374
35
<220>
<221> sig_peptide
<222> 222..299
<223> Von Heijne matrix
      score 4.28353322771141
40      seq ILFKFSLCPYAAA/LS

<400> 140
taataatggt gttaaattat tgccttctca tctgcgtgtc tcttatgttc tgcttaaaga      60
gattgtcagt ttgttcaagc tctttttagt tgttgtctct ccagtgccta gctttgagct      120
45  ttgtacacgg tagttattga gttgagtaac atagtttggt ctgagtcatt tgttccacat      180
gcttgaagac ttggcttaac ctagtagata ataggaaaga a atg gaa atg ctc ttt      236
      Met Glu Met Leu Phe
      -25
50  gat gaa aga gcc cct ctc tta ttc atc ctt ttt aaa ttt tct ttg tgc      284
Asp Glu Arg Ala Pro Leu Leu Phe Ile Leu Phe Lys Phe Ser Leu Cys
      -20              -15              -10
cca tat gca gca gct ctc agc aaa cct ata ttt ggc agt gtg gcc tgt      332
Pro Tyr Ala Ala Ala Leu Ser Lys Pro Ile Phe Gly Ser Val Ala Cys
      -5              1              5              10
55  atg act aaa gaa atc ctg gcc agg cac ggt ggc tca cgc ctg      374
Met Thr Lys Glu Ile Leu Ala Arg His Gly Gly Ser Arg Leu
      15              20              25
taatcccagc actttgggag gccgaggcgg gtggattacg aggtcaggag attgagacca      434
tcttggttaa catggcgaaa ccccatctct acgaaaaata caaaaaaaaaa aattagccgg      494
60  gcatcatggc gggcgctgt agtcttagct actcaggagg ctgaggcagg agaatggcgt      554
gaaccgggga ggcggagctt gcagttagcc gagattgcgc cactgcactc cagcctgggg      614
caacagagca agactccgtc tcaaaaaaaaa aaaaaaaaaa      653

```

```

<210> 141
<211> 1490
<212> DNA
<213> Homo sapiens
5
<220>
<221> CDS
<222> 59..274

10 <220>
<221> sig_peptide
<222> 59..127
<223> Von Heijne matrix
      score 7.37647149292058
15      seq LGLCSLLVGEAEA/PS

<400> 141
agacagaggc agggccttgcg acggaagtgg cctctctgct tctgcagggc tggggaag      58
atg ctg cgt cca gcg tta ccg tgg ctg tac ctt ggc ctc tgc agc ctc      106
20 Met Leu Arg Pro Ala Leu Pro Trp Leu Tyr Leu Gly Leu Cys Ser Leu
      -20 -15 -10
ctg gtg ggg gag gca gag gcc ccg agc ccc gtg gat ccg ctg gag cgg      154
Leu Val Gly Glu Ala Glu Ala Pro Ser Pro Val Asp Pro Leu Glu Arg
      -5 1 5
25 agc cgg ccg tac gcg gtg ctg cga ggg cag aac ctg gtg ttg atg gga      202
Ser Arg Pro Tyr Ala Val Leu Arg Gly Gln Asn Leu Val Leu Met Gly
      10 15 20 25
acc att ttc agc atc ctg ctg gtg act gtc atc ctt atg gca ttt tgt      250
Thr Ile Phe Ser Ile Leu Leu Val Thr Val Ile Leu Met Ala Phe Cys
30      30 35 40
gtc tac aag ccc att ccg cgt ccg tgacagccag acaagttctt caatgagtat      304
Val Tyr Lys Pro Ile Arg Arg Arg
      45
ttgggaatag gataagttgt gttgcacaca ggccagtggg gaagttggaa ccaaaacttt      364
35 cctacttgga aatgaccttt ggtctggaca gttggtaaat gctaaatgaa ttagaagaaa      424
acatgtacta gacattatth tttcctaaca ctgtagcgca aataattggc cctgagtc      484
gcttctcagt gtttctgact gtacttggtta aaagtaagac ctgaaagctc caaaggtcag      544
tgtaaagatg gagtgttcat gagaaagaaa acatggtaac cttgtgagtg cctgtaagaa      604
ccacactgta aagaactcat cattaatgct tgaaaatggt attaagaagg agacttacca      664
40 tgcagacatt ccctatttaa gaaccatttg gttacagtgg gttaagaatc acagattttt      724
ttttttaatc tcacctgagt tagcctagaa tgcgctgggt gcaaagtggg gtcagctgtg      784
gggatcttgg gccctcggtc ctcacctgca tcctgccctg cactcagggt cccccctga      844
agtcagggtc acatcaggta gacctgttac tatatgcacc tttggcctgg aatgctctga      904
agttggactg gaaatgttac taggttggcc tgttacaaaa aggaccccat cctgcttaaa      964
45 cacattgac tcccttgccc tgcatttgag tctttctagc ccacggtctg aaacttgagg      1024
cagctttcca gatttggaat gtaaaaggct cagtgggcac tctgttcac cctgggtggg      1084
gagggccag ccaacagaag tgcatgtcca ctgtgcgggc cagtgtgtgt ttacacaaat      1144
ttcatctcag ctttgaaaat gctgctatta gtttccactg ttggtgaact ggattttttc      1204
ctcctattga aatgatactt tcatacttat aaagctgtcg tcaatattta tttcaagggt      1264
50 ctagatttaa ttttggtatt aaattgaaat gcttatcttg tgttcaagca cagcactgat      1324
tttaacaacc tgcatttaat gtgaagtaac cgaagtagga tactgtaact gtgtaaggat      1384
tttgtttgta atcttgtaac attgaaccat tgaaatgttc agttctttgc ttttgagcaa      1444
aacgtcaatt aaaactaaag taaaatctta aaaaaaaaaa aaaaaa      1490

55 <210> 142
<211> 661
<212> DNA
<213> Homo sapiens

60 <220>
<221> CDS
<222> 158..442

```

<220>
 <221> sig_peptide
 <222> 158..301
 <223> Von Heijne matrix
 5 score 7.53908709538105
 seq FVILLLFIFTVVS/LV

<400> 142
 10 aaaaacagac gataccatcg cttcagcagc atcctctcag acaagagcca ctatttctga 60
 ttcagatcac ctgtcatcga agtttaaaga aggggaaaca ggagacagaa atacactgaa 120
 ccaaaaagat tcaaaagagc aagtggaatc tctaaga atg gct tcc agc cac tgg 175
 Met Ala Ser Ser His Trp
 -45
 15 aat gaa acc act acc tct gtt tat cag tac ctt ggt ttt caa gtt caa 223
 Asn Glu Thr Thr Thr Ser Val Tyr Gln Tyr Leu Gly Phe Gln Val Gln
 -40 -35 -30
 aaa att tac cct ttc cat gac aac tgg aac act gcc tgc ttt gtc atc 271
 Lys Ile Tyr Pro Phe His Asp Asn Trp Asn Thr Ala Cys Phe Val Ile
 -25 -20 -15
 20 ctg ctt tta ttt ata ttt aca gtg gta tct tta gtg gtg ctg gct ttc 319
 Leu Leu Leu Phe Ile Phe Thr Val Val Ser Leu Val Val Leu Ala Phe
 -10 -5 1 5
 ctt tat gaa gtg ctt gac tgc tgc tgc tgt gta aaa aac aaa acc gtg 367
 Leu Tyr Glu Val Leu Asp Cys Cys Cys Cys Val Lys Asn Lys Thr Val
 25 10 15 20
 aaa gac ttg aaa agt gaa ccc aac cct ctt aga agt atg atg gac aac 415
 Lys Asp Leu Lys Ser Glu Pro Asn Pro Leu Arg Ser Met Met Asp Asn
 25 30 35
 atc aga aaa cgt gaa act gaa gtg gtc taacactcta tagaagatga 462
 30 Ile Arg Lys Arg Glu Thr Glu Val Val
 40 45
 acaaaatctc tgaaagcagc tcaacctctt ctgagaaaaa aaatatattc tgaggccaac 522
 tgttgctaca aaacaaattc tgactgaatg tttaaaacat ttctagtaga aggggaaaaa 582
 aaagttaaac atgcactgtt tgtgtgtata gccatttcat taaatatata gtaaaacttc 642
 35 ataaaaaaaa aaaaaaaaaa 661

<210> 143
 <211> 1789
 <212> DNA
 40 <213> Homo sapiens

<220>
 <221> CDS
 <222> 5..454
 45

<220>
 <221> sig_peptide
 <222> 5..64
 <223> Von Heijne matrix
 50 score 6.64507667657896
 seq LLPLLSLLVGAWL/KL

<400> 143
 55 cctg atg gcc cgg cat ggg tta ccg ctg ctg ccc ctg ctg tgc ctc ctg 49
 Met Ala Arg His Gly Leu Pro Leu Leu Pro Leu Leu Ser Leu Leu
 -20 -15 -10
 gtc ggc gcg tgg ctc aag cta gga aat gga cag gct act agc atg gtc 97
 Val Gly Ala Trp Leu Lys Leu Gly Asn Gly Gln Ala Thr Ser Met Val
 -5 1 5 10
 60 caa ctg cag ggt ggg aga ttc ctg atg gga aca aat tct cca gac agc 145
 Gln Leu Gln Gly Gly Arg Phe Leu Met Gly Thr Asn Ser Pro Asp Ser
 15 20 25
 aga gat ggt gaa ggg cct gtg cgg gag gcg aca gtg aaa ccc ttt gcc 193

	Arg	Asp	Gly	Glu	Gly	Pro	Val	Arg	Glu	Ala	Thr	Val	Lys	Pro	Phe	Ala	
			30					35					40				
	atc	gac	ata	ttt	cct	gtc	acc	aac	aaa	gat	ttc	agg	gat	ttt	gtc	agg	241
	Ile	Asp	Ile	Phe	Pro	Val	Thr	Asn	Lys	Asp	Phe	Arg	Asp	Phe	Val	Arg	
5		45					50					55					
	gag	aaa	aag	tat	cgg	aca	gaa	gct	gag	atg	ttt	gga	tgg	agc	ttt	gtc	289
	Glu	Lys	Lys	Tyr	Arg	Thr	Glu	Ala	Glu	Met	Phe	Gly	Trp	Ser	Phe	Val	
	60					65					70				75		
	ttt	gag	gac	ttt	gtc	tct	gat	gag	ctg	aga	aac	aaa	gcc	acc	cag	cca	337
10	Phe	Glu	Asp	Phe	Val	Ser	Asp	Glu	Leu	Arg	Asn	Lys	Ala	Thr	Gln	Pro	
				80						85					90		
	atg	aag	gtc	aag	ttt	acc	cat	ggg	gga	act	ggg	tcc	agc	caa	acc	gca	385
	Met	Lys	Val	Lys	Phe	Thr	His	Gly	Gly	Thr	Gly	Ser	Ser	Gln	Thr	Ala	
				95					100					105			
15	cca	acc	tgt	ggc	agg	gaa	agt	tcc	cca	agg	gag	aca	aag	ctg	agg	atg	433
	Pro	Thr	Cys	Gly	Arg	Glu	Ser	Ser	Pro	Arg	Glu	Thr	Lys	Leu	Arg	Met	
			110					115					120				
	gct	tcc	atg	gag	tct	ccc	cag	tgaatgcttt	ccccgcccag	aacaactacg							484
	Ala	Ser	Met	Glu	Ser	Pro	Gln										
20		125					130										
	ggctctatga	cctcctgggg	aacgtgtggg	agtggacagc	atcacccgtac	caggctgctg											544
	agcaggacat	gcgcgtcctc	cggggggcat	cctggatcga	cacagctgat	ggctctgccca											604
	atcacccgggc	ccgggtcacc	accaggatgg	gcaacactcc	agattcagcc	tcagacaacc											664
	tcggtttccg	ctgtgctgca	gacgcaggcc	ggccgccagg	ggagctgtaa	gcagccgggt											724
25	ggtgacaagg	agaaaagcct	tctaggggtca	ctgtcattcc	ctggccatgt	tgcaaacagc											784
	gcaattccaa	gctcgagagc	ttcagcctca	ggaaagaact	tccccctccc	tgtctcccat											844
	ccctctgtgg	caggcgccctc	tcaccagggc	aggagaggac	tcagccctcc	gtgttttgga											904
	gaagggggccc	aatgtgtgtt	gacgatggct	ggggggccagg	tggttctgtt	agaggccaag											964
	tattattgac	acaggattgc	aaacacacaa	acaattggaa	cagagcactc	tgaaaggcca											1024
30	ttttttaagc	attttaaaat	ctattctctc	cccccttctc	cctggatgat	tcaggaagct											1084
	gacattgttt	cctcaaggca	gaattttcct	ggttctgttt	tctcagccag	ttgctgtgga											1144
	aggagaatgc	tttctttgtg	gcctcatctg	tggtttcgtg	tccctctgaa	ggaaactagt											1204
	ttccactgtg	taacaggcag	acatgttaact	atttaaagca	cagttcagtc	ctaaaagggt											1264
	ctggggagaac	cagatgatgt	actagggtgaa	gcattgcatt	gtgggaatca	caaagcaaact											1324
35	agtactccag	aaagacaaaat	atcagaagct	tcctattctt	tttttttttt	tttttttttt											1384
	ttgagacagg	gtctttctct	gttgcccagg	ctagagtgc	ctggtgatca	cggctcactc											1444
	tagccttgaa	ttcctggggc	caagcaattc	tcccactca	gcctcctgag	tagctgggac											1504
	tacaagtgtg	caccaccatg	cctggctaact	tttttgaaat	tttgtagtga	tgggatctcg											1564
	ctctgttgcc	cagggtgggc	tcgaactcct	ggcctcaagc	gacccctcca	cctcgacctc											1624
40	ccaaagtgtc	gggattacag	gtgtgagcca	cctcgctgg	gcccccttct	ccatagtcc											1684
	ccaaaaacat	gtccctggag	agtagcctgc	tcccacactg	tcactggatg	tcatggggcc											1744
	aataaaatct	cctgcaattg	tgtatctcaa	aaaaaaaaaa	aaaaa												1789
	<210> 144																
45	<211> 2006																
	<212> DNA																
	<213> Homo sapiens																
	<220>																
50	<221> CDS																
	<222> 241..1302																
	<400> 144																
	tagtgccgga	gccccgccag	agccccgactt	cagccccagc	cagatcccg	gtcaacggag											60
55	gcggaacggc	ggacccccgta	ccctggcagc	atcgagacac	cggcggtga	aggcaagggtc											120
	cctggactgg	tcataacact	cttgtggccc	tggcagaatc	aagatgaggc	cctgtcatgc											180
	ctccccagtg	aggcctacag	tctgagcaga	cagcatggcc	tgccactggc	agtgaacacc											240
	atg tct gca gga ggt ggc cgg gcc ttt gct tgg	caa	gtg	ttc	ccc	ccc											288
	Met Ser Ala Gly Gly Gly Arg Ala Phe Ala Trp	Gln	Val	Phe	Pro	Pro											
60	1		5				10					15					
	atg	ccc	act	tgc	cgg	gtc	tat	ggc	aca	gtg	gca	cac	caa	gat	ggg	cac	336
	Met	Pro	Thr	Cys	Arg	Val	Tyr	Gly	Thr	Val	Ala	His	Gln	Asp	Gly	His	
				20					25					30			

	ctg	ctg	gtg	ttg	ggg	ggt	tgt	ggc	cgg	gct	gga	ctg	ccc	ctg	gac	act	384
	Leu	Leu	Val	Leu	Gly	Gly	Cys	Gly	Arg	Ala	Gly	Leu	Pro	Leu	Asp	Thr	
			35					40					45				
5	gct	gag	aca	ctg	gac	atg	gcc	tcg	cac	aca	tgg	ctg	gca	ctg	gca	ccc	432
	Ala	Glu	Thr	Leu	Asp	Met	Ala	Ser	His	Thr	Trp	Leu	Ala	Leu	Ala	Pro	
		50					55					60					
	ctg	ccc	act	gcc	cgg	gct	ggt	gca	gct	gcg	gta	gtt	ctg	ggc	aag	cag	480
	Leu	Pro	Thr	Ala	Arg	Ala	Gly	Ala	Ala	Ala	Val	Val	Leu	Gly	Lys	Gln	
		65				70					75					80	
10	gtg	cta	gtg	gtg	tgt	ggt	gtg	gat	gag	gtc	cag	agc	ccg	gta	gct	gct	528
	Val	Leu	Val	Val	Cys	Gly	Val	Asp	Glu	Val	Gln	Ser	Pro	Val	Ala	Ala	
					85					90					95		
	gta	gag	gcc	ttc	ctg	atg	gat	gag	ggc	cgc	tgg	gag	cgt	cgg	gcc	acc	576
	Val	Glu	Ala	Phe	Leu	Met	Asp	Glu	Gly	Arg	Trp	Glu	Arg	Arg	Ala	Thr	
15				100					105					110			
	ctc	cct	caa	gca	gcc	atg	ggg	gtt	gca	act	gtg	gag	aga	gat	ggt	atg	624
	Leu	Pro	Gln	Ala	Ala	Met	Gly	Val	Ala	Thr	Val	Glu	Arg	Asp	Gly	Met	
			115				120						125				
20	gtg	tat	gct	ctg	ggg	gga	atg	ggc	cct	gac	acg	gcc	ccc	cag	gcc	cag	672
	Val	Tyr	Ala	Leu	Gly	Gly	Met	Gly	Pro	Asp	Thr	Ala	Pro	Gln	Ala	Gln	
		130					135						140				
	gta	cgt	gtg	tat	gac	ccc	cgt	cgg	gac	tgc	tgg	ctt	tcg	cta	ccc	tcc	720
	Val	Arg	Val	Tyr	Asp	Pro	Arg	Arg	Asp	Cys	Trp	Leu	Ser	Leu	Pro	Ser	
		145				150					155					160	
25	atg	ccc	aca	ccc	tgc	tat	ggg	gcc	tcc	acc	ttc	ctg	cac	ggg	aac	aag	768
	Met	Pro	Thr	Pro	Cys	Tyr	Gly	Ala	Ser	Thr	Phe	Leu	His	Gly	Asn	Lys	
				165			170							175			
	atc	tat	gtc	ctg	ggg	ggc	cgc	cag	ggc	aag	ctc	ccg	gtg	act	gct	ttt	816
	Ile	Tyr	Val	Leu	Gly	Gly	Arg	Gln	Lys	Lys	Leu	Pro	Val	Thr	Ala	Phe	
30				180					185					190			
	gaa	gcc	ttt	gat	ctg	gag	gcc	cgt	aca	tgg	acc	cgg	cat	cca	agc	cta	864
	Glu	Ala	Phe	Asp	Leu	Glu	Ala	Arg	Thr	Trp	Thr	Arg	His	Pro	Ser	Leu	
			195				200					205					
35	ccc	agc	cgt	cgg	gcc	ttt	gct	ggc	tgc	gcc	atg	gct	gaa	ggc	agc	gtc	912
	Pro	Ser	Arg	Arg	Ala	Phe	Ala	Gly	Cys	Ala	Met	Ala	Glu	Gly	Ser	Val	
		210					215					220					
	ttt	agc	ctg	ggt	ggc	ctg	cag	cag	cct	ggg	ccc	cac	aac	ttc	tac	tct	960
	Phe	Ser	Leu	Gly	Gly	Leu	Gln	Gln	Pro	Gly	Pro	His	Asn	Phe	Tyr	Ser	
		225				230					235					240	
40	cgc	cca	cac	ttt	gtc	aac	act	gtg	gag	atg	ttt	gac	ctg	gag	cat	ggg	1008
	Arg	Pro	His	Phe	Val	Asn	Thr	Val	Glu	Met	Phe	Asp	Leu	Glu	His	Gly	
				245						250					255		
	tcc	tgg	acc	aaa	ttg	ccc	cgc	agc	ctg	cgc	atg	agg	gat	aag	agg	gca	1056
	Ser	Trp	Thr	Lys	Leu	Pro	Arg	Ser	Leu	Arg	Met	Arg	Asp	Lys	Arg	Ala	
45				260					265					270			
	gac	ttt	gtg	gtt	ggg	tcc	ctt	ggg	ggc	cac	att	gtg	gcc	att	ggg	ggc	1104
	Asp	Phe	Val	Val	Gly	Ser	Leu	Gly	Gly	His	Ile	Val	Ala	Ile	Gly	Gly	
			275				280						285				
	ctt	gga	aac	cag	cca	tgt	cct	ttg	ggc	tct	gtg	gag	agc	ttt	agc	ctt	1152
50	Leu	Gly	Asn	Gln	Pro	Cys	Pro	Leu	Gly	Ser	Val	Glu	Ser	Phe	Ser	Leu	
		290					295					300					
	gca	cgg	cgg	cgc	tgg	gag	gca	ttg	cct	gcc	atg	ccc	act	gcc	cgc	tgc	1200
	Ala	Arg	Arg	Arg	Trp	Glu	Ala	Leu	Pro	Ala	Met	Pro	Thr	Ala	Arg	Cys	
		305				310					315					320	
55	tcc	tgc	tct	agt	ctg	cag	gct	ggg	ccc	cgg	ctg	ttt	gtt	att	ggg	ggt	1248
	Ser	Cys	Ser	Ser	Leu	Gln	Ala	Gly	Pro	Arg	Leu	Phe	Val	Ile	Gly	Gly	
					325				330						335		
	gtg	gcc	cag	ggc	ccc	agt	caa	gcc	gtg	gag	gca	ctg	tgt	ctg	cgt	gat	1296
	Val	Ala	Gln	Gly	Pro	Ser	Gln	Ala	Val	Glu	Ala	Leu	Cys	Leu	Arg	Asp	
60				340					345					350			
	ggg	gtc	tga	agg	cttg	gtggg	agctg	tccact	ggag	cagctc	attg	ccagagg	cag				1352
	Gly	Val															
	ctattttctat	ggctcctttt	gctgctgagg	acactcactg	tggctctgtg	ggatgagaga											1412

ggcattggggg tgagcacttg aaacactgcc ttggggcctt gggttagggg agcctttgtc 1472
 ttttagtgac gacacacata tgcttacacc tacctttatc accattcggt catgaatcat 1532
 gcctagctcc atccttgccc tgggacctac taggccttcc atccaactgg gaaatgggga 1592
 gaagcaaaagc tggcctcatg ctcttcagggt tcagttccta tctggagttg accaggccta 1652
 5 cccagattgc cattcctgaa aaatctcagc tgccaggctg cctttagggg ccctgcagac 1712
 ccaggagagt tgagaggggt ggggacacac acagaataga gaggatgtgg gaactgccag 1772
 agggccggag cgcaggaggt caagtggagg aatgctgggt ttgagccctc tacactgctg 1832
 gttgtatgac cttggacaag tcacttcacc tctctgtgcc tcagcatcct catctataaa 1892
 tggggatctc tgaaaccttc ctaccctacc tacctcacag ggctgtttgtg aggacccagg 1952
 10 gaggtttgat gtggaagtaa aagtgtgtgt aaaaccgaaa aaaaaaaaaa aaaa 2006

<210> 145

<211> 1096

<212> DNA

15 <213> Homo sapiens

<220>

<221> CDS

<222> 15..635

20

<400> 145

atccaaggcg caag atg gcg ctg ctt ttt gca cgt tct ttg cgc ttg tgc 50
 Met Ala Leu Leu Phe Ala Arg Ser Leu Arg Leu Cys
 1 5 10
 25 cgc tgg gga gcc aaa cga ttg gga gtt gcc tcc aca gag gcc cag aga 98
 Arg Trp Gly Ala Lys Arg Leu Gly Val Ala Ser Thr Glu Ala Gln Arg
 15 20 25
 ggc gtc agt ttc aaa ctg gaa gaa aaa acc gcc cac agc agc ctg gca 146
 Gly Val Ser Phe Lys Leu Glu Glu Lys Thr Ala His Ser Ser Leu Ala
 30 30 35 40
 ctc ttc aga gat gat acg ggt gtc aaa tat ggc ttg gtg gga ttg gag 194
 Leu Phe Arg Asp Asp Thr Gly Val Lys Tyr Gly Leu Val Gly Leu Glu
 45 50 55 60
 ccc acc aag gtg gcc ttg aat gtg gag cgc ttc cgg gag tgg gca gtg 242
 Pro Thr Lys Val Ala Leu Asn Val Glu Arg Phe Arg Glu Trp Ala Val
 65 70 75
 gtg ctg gca gac aca gcg gtc acc agt ggc aga cac tac tgg gaa gtg 290
 Val Leu Ala Asp Thr Ala Val Thr Ser Gly Arg His Tyr Trp Glu Val
 80 85 90
 40 aca gtg aag cgc tcc cag cag ttc cgg ata gga gtg gca gat gtg gac 338
 Thr Val Lys Arg Ser Gln Gln Phe Arg Ile Gly Val Ala Asp Val Asp
 95 100 105
 atg tcc cgg gat agc tgc att ggt gtt gat gat cgt tcc tgg gtg ttc 386
 Met Ser Arg Asp Ser Cys Ile Gly Val Asp Asp Arg Ser Trp Val Phe
 110 115 120
 45 acc tat gcc cag cgc aag tgg tac acc atg ttg gcc aac gag aaa gcc 434
 Thr Tyr Ala Gln Arg Lys Trp Tyr Thr Met Leu Ala Asn Glu Lys Ala
 125 130 135 140
 cca gtt gag ggt att ggg cag cca gag aag gtg ggg ctg ttg ctg gag 482
 50 Pro Val Glu Gly Ile Gly Gln Pro Glu Lys Val Gly Leu Leu Leu Glu
 145 150 155
 tat gag gcc cag aag ctg agc ctg gtg gat gtg agc cag gtc tct gtg 530
 Tyr Glu Ala Gln Lys Leu Ser Leu Val Asp Val Ser Gln Val Ser Val
 160 165 170
 55 gtt cac acg cta cag aca gat ttc cgg ggt cca gtg gtg cct gcc ttt 578
 Val His Thr Leu Gln Thr Asp Phe Arg Gly Pro Val Val Pro Ala Phe
 175 180 185
 gct ctc tgg gat ggg gag ctg ctg acc cat tca ggg ctt gag gtg ccc 626
 Ala Leu Trp Asp Gly Glu Leu Leu Thr His Ser Gly Leu Glu Val Pro
 190 195 200
 60 gag ggc ctc tagtatgtcc attactggag tccctaataca cgcctttggc 675
 Glu Gly Leu
 205

	cagcctcctt	ttgaaagtgt	ccgaagcctt	tttactttgc	ctcaagcaac	ctctagctcc	735
	cacaattcag	tgttggtgcc	tctgtgcaat	atcatgatca	tcttcctcat	cccctacctt	795
	gtgaaagcta	ggcatacagc	caaaccctcc	ttttccccac	ccaccaaacac	tactgccaat	855
	ttcctaggct	accatgggtg	tatcttcctt	gacctgcttc	cttcagtccc	tctgctcccc	915
5	tttgcccagg	cctttctcag	actgtattcc	atcctggggt	cttatcattc	agctttgttt	975
	gaatttatta	atcaccatga	tacctctccc	tccctttgtc	cacatgtaac	ttgttcttgg	1035
	ggctctacca	gatggctgaa	gagtaaatacc	tttctacctc	tggcaaaaaa	aaaaaaaaaa	1095
	a						1096
10	<210>	146					
	<211>	1666					
	<212>	DNA					
	<213>	Homo sapiens					
15	<220>						
	<221>	CDS					
	<222>	109..738					
	<400>	146					
20	cccagcgcttc	ctcctccggc	cccaggtcac	cgccagcacg	cgcttgcttc	cggtctgcgc	60
	gagtccacgc	agctccccag	gcccttcacc	agcacagcag	cagcaggc	atg gca gca	117
						Met Ala Ala	
						1	
25	agc gtg gag cag cgc gag ggc acc atc cag gtg cag ggc cag gcc ctc						165
	Ser Val Glu Gln Arg Glu Gly Thr Ile Gln Val Gln Gly Gln Ala Leu						
	5 10 15						
	ttc ttc cga gag gcc ctg ccc ggc agt ggg cag gct cgc ttc tct gta						213
	Phe Phe Arg Glu Ala Leu Pro Gly Ser Gly Gln Ala Arg Phe Ser Val						
	20 25 30 35						
30	ctg ctg ctg cat ggt att cgc ttc tcc tcc gag acc tgg cag aac ctg						261
	Leu Leu Leu His Gly Ile Arg Phe Ser Ser Glu Thr Trp Gln Asn Leu						
	40 45 50						
	ggt aca ctg cac agg ctg gcc cag gct ggc tac cgg gct gtg gcc att						309
	Gly Thr Leu His Arg Leu Ala Gln Ala Gly Tyr Arg Ala Val Ala Ile						
35	55 60 65						
	gac ctg cca ggt ctg ggg cac tcc aag gaa gca gca gcc cct gcc cct						357
	Asp Leu Pro Gly Leu Gly His Ser Lys Glu Ala Ala Pro Ala Pro						
	70 75 80						
40	att ggg gag ctg gcc cct ggc agc ttc ctg gcg gct gtg gtg gat gcc						405
	Ile Gly Glu Leu Ala Pro Gly Ser Phe Leu Ala Ala Val Val Asp Ala						
	85 90 95						
	ttg gag ctg ggc ccc ccg gtt gtg atc agt cca tca ctg agt ggc atg						453
	Leu Glu Leu Gly Pro Pro Val Val Ile Ser Pro Ser Leu Ser Gly Met						
	100 105 110 115						
45	tac tcc ctg ccc ttc ctc acg gcc cct ggc tcc cag ctc ccg ggc ttt						501
	Tyr Ser Leu Pro Phe Leu Thr Ala Pro Gly Ser Gln Leu Pro Gly Phe						
	120 125 130						
	gtg cca gtg gcc ccc atc tgc act gac aaa atc aat gct gcc aac tat						549
	Val Pro Val Ala Pro Ile Cys Thr Asp Lys Ile Asn Ala Ala Asn Tyr						
50	135 140 145						
	gcc agt gtg aag act cca gct ctg att gta tat gga gac cag gac ccc						597
	Ala Ser Val Lys Thr Pro Ala Leu Ile Val Tyr Gly Asp Gln Asp Pro						
	150 155 160						
	atg ggt cag acc agc ttt gag cac ctg aag cag ctg ccc aac cac cgg						645
55	Met Gly Gln Thr Ser Phe Glu His Leu Lys Gln Leu Pro Asn His Arg						
	165 170 175						
	gtg ctg atc atg aag ggg gcg ggg cac ccc tgt tac ctg gac aaa cca						693
	Val Leu Ile Met Lys Gly Ala Gly His Pro Cys Tyr Leu Asp Lys Pro						
	180 185 190 195						
60	gag gag tgg cat aca ggg ctg ctg gac ttc ctg cag ggg ctc cag						738
	Glu Glu Trp His Thr Gly Leu Leu Asp Phe Leu Gln Gly Leu Gln						
	200 205 210						
	tgaagcccag cactgctgca ggggggtgggc tgctgctg ctctgagctc tctcttgca						798

```

gctctctctt ctctcccagg ctctgggtca tgcacatgca acaggtgcgt ctgtctatat 858
gtctgggttc ttgtcttttg tggctgtgtt gtcttttcta cctctttctc ttgcagtgat 918
agactgaggg ggtaaaatca agagaaaaaa ctctcaggaa tcaaggaaca taatcctgtg 978
gagggtaatc cattacatga gcttctcctg ttcttccact ttcttgctg gctttcactc 1038
5 cttccccctgc tctgcccagc ctttccctcc caccactcc tacttctgca aatgccctga 1098
aggccagccc ttaccccaac acccacttcc ccacctcctt aggccccaga tacatacatg 1158
cccacatgca cgcttacatg tttagagcca tccttgtttc caaatatgac ccttcgcttg 1218
agggcaactg cataggtaca tctaactctg gactggcatg cacattgtca tgtgcagctt 1278
tgcatataca cacatgcata catgagcctc cacacaagca cttgcacaca tgtggactcc 1338
10 taaccatgct aacctcactg gctgggaagg tggggacccc atggggccagc ccttgccagga 1398
ggcccttttg caaggcttag ggtgtggcca gccctgaaag ctacttggac acagggttca 1458
gctggcccca gccagaagt gacccccaga aagggagggc caccgctttg cccctgctt 1518
ttacccttcc ttctgggtgc tctacacctc aggttaccag gcctgaggca tctcagccaa 1578
gcttggtttcc tgctctgagg cttgtggggt gggagccaga gtggagggtcg gtgaaataaa 1638
15 gtgatgcaat taaaaaaaaa aaaaaaaaaa 1666

```

<210> 147

<211> 1687

<212> DNA

20 <213> Homo sapiens

<220>

<221> CDS

<222> 21..1145

25

<400> 147

```

gtttctccgg acttcgagcc atg gcg gtg acg gaa gcg agc ctg ttg cgc cag 53
Met Ala Val Thr Glu Ala Ser Leu Leu Arg Gln
1 5 10
30 tgc ccc ctg ctt ctg ccc cag aac cgg tgc aaa acc gtg tat gag gga 101
Cys Pro Leu Leu Leu Pro Gln Asn Arg Ser Lys Thr Val Tyr Glu Gly
15 20 25
ttc atc tcg gct cag gga aga gac ttc cac ctt agg ata gtg ttg cct 149
Phe Ile Ser Ala Gln Gly Arg Asp Phe His Leu Arg Ile Val Leu Pro
30 35 40
35 gaa gat tta caa ctg aag aat gca aga tta tta tgt att tgg cag ctg 197
Glu Asp Leu Gln Leu Lys Asn Ala Arg Leu Leu Cys Ile Trp Gln Leu
45 50 55
aga aca ata ctt agt gga tac cat cga ata gta caa cag aga atg cag 245
Arg Thr Ile Leu Ser Gly Tyr His Arg Ile Val Gln Gln Arg Met Gln
60 65 70 75
cac tct cct gat cta atg agc ttt atg atg gag ttg aag atg ctt ttg 293
His Ser Pro Asp Leu Met Ser Phe Met Met Glu Leu Lys Met Leu Leu
80 85 90
45 gaa gtt gcc tta aag aat aga caa gag ctg tat gca cta cct cct cct 341
Glu Val Ala Leu Lys Asn Arg Gln Glu Leu Tyr Ala Leu Pro Pro Pro
95 100 105
ccc cag ttc tac tca agc ctt att gaa gag ata gga act ctt ggt tgg 389
Pro Gln Phe Tyr Ser Ser Leu Ile Glu Glu Ile Gly Thr Leu Gly Trp
110 115 120
50 gat aaa ctt gtg tat gcg gat acc tgc ttc agt acc atc aag tta aaa 437
Asp Lys Leu Val Tyr Ala Asp Thr Cys Phe Ser Thr Ile Lys Leu Lys
125 130 135
gca gaa gat gct tct ggt aga gag cat tta atc act ctc aag ttg aag 485
Ala Glu Asp Ala Ser Gly Arg Glu His Leu Ile Thr Leu Lys Leu Lys
140 145 150 155
gca aag tat cct gca gaa tca cca gat tat ttt gtg gat ttt cct gtt 533
Ala Lys Tyr Pro Ala Glu Ser Pro Asp Tyr Phe Val Asp Phe Pro Val
160 165 170
60 cca ttt tgt gcc tcc tgg aca cct cag agc tcc tta ata agc att tat 581
Pro Phe Cys Ala Ser Trp Thr Pro Gln Ser Ser Leu Ile Ser Ile Tyr
175 180 185
agt cag ttt ttg gca gca ata gaa tca cta aag gca ttc tgg gat gtt 629

```

	Ser	Gln	Phe	Leu	Ala	Ala	Ile	Glu	Ser	Leu	Lys	Ala	Phe	Trp	Asp	Val	
		190						195					200				
	atg	gat	gaa	atc	gat	gag	aag	acc	tgg	gta	ctt	gag	cca	gaa	aaa	cct	677
	Met	Asp	Glu	Ile	Asp	Glu	Lys	Thr	Trp	Val	Leu	Glu	Pro	Glu	Lys	Pro	
5		205					210					215					
	cca	cgg	agt	gca	aca	gca	cgc	aga	att	gca	tta	ggg	aat	aat	gtt	tcc	725
	Pro	Arg	Ser	Ala	Thr	Ala	Arg	Arg	Ile	Ala	Leu	Gly	Asn	Asn	Val	Ser	
		220				225					230				235		
	ata	aat	ata	gag	gta	gac	ccc	agg	cat	cct	act	atg	ctt	cct	gag	tgc	773
10	Ile	Asn	Ile	Glu	Val	Asp	Pro	Arg	His	Pro	Thr	Met	Leu	Pro	Glu	Cys	
					240						245				250		
	ttc	ttt	ctt	gga	gct	gac	cat	gtg	gta	aaa	ccc	ctg	gga	att	aag	ctg	821
	Phe	Phe	Leu	Gly	Ala	Asp	His	Val	Val	Lys	Pro	Leu	Gly	Ile	Lys	Leu	
				255					260				265				
15	agc	agg	aac	ata	cat	ttg	tgg	gat	cca	gaa	aat	agt	gtg	tta	caa	aat	869
	Ser	Arg	Asn	Ile	His	Leu	Trp	Asp	Pro	Glu	Asn	Ser	Val	Leu	Gln	Asn	
			270					275				280					
	ttg	aaa	gat	gtt	tta	gaa	att	gat	ttt	cca	gct	cgt	gct	atc	ctg	gaa	917
	Leu	Lys	Asp	Val	Leu	Glu	Ile	Asp	Phe	Pro	Ala	Arg	Ala	Ile	Leu	Glu	
20		285					290					295					
	aaa	tct	gat	ttt	act	atg	gat	tgt	gga	att	tgt	tat	gct	tat	caa	ctt	965
	Lys	Ser	Asp	Phe	Thr	Met	Asp	Cys	Gly	Ile	Cys	Tyr	Ala	Tyr	Gln	Leu	
		300				305					310				315		
	gac	ggg	acc	att	cct	gat	caa	gtg	tgt	gat	aat	tcc	cag	tgt	gga	caa	1013
25	Asp	Gly	Thr	Ile	Pro	Asp	Gln	Val	Cys	Asp	Asn	Ser	Gln	Cys	Gly	Gln	
				320					325				330				
	cct	ttc	cat	caa	ata	tgc	tta	tat	gag	tgg	ctg	aga	gga	cta	cta	act	1061
	Pro	Phe	His	Gln	Ile	Cys	Leu	Tyr	Glu	Trp	Leu	Arg	Gly	Leu	Leu	Thr	
				335					340				345				
30	agt	aga	cag	agt	ttt	aac	atc	ata	ttt	ggg	gaa	tgt	cca	tat	tgt	agt	1109
	Ser	Arg	Gln	Ser	Phe	Asn	Ile	Ile	Phe	Gly	Glu	Cys	Pro	Tyr	Cys	Ser	
			350				355					360					
	aag	cca	att	acc	tta	aaa	atg	tct	gga	agg	aaa	cac	tgaaataaga				1155
	Lys	Pro	Ile	Thr	Leu	Lys	Met	Ser	Gly	Arg	Lys	His					
35		365					370					375					
	atacaacatt	tcgggtgaaga	gctggaaact	taaaaaatta	tcaaaaggaa	ttttggtatc											1215
	atcttcagag	aaaaaataaa	gcaagraata	ctaactcaa	aaggacagg	atgatgatgc											1275
	gataataata	aacatctgcg	tttgtctctt	cactaagagt	aaactgggaa	attgtaggcc											1335
	aaagtccagt	tgaactttct	aagtctgtga	tccccgtgct	gactgtggaa	gtgtatttat											1395
40	accaagatgg	agatcttgac	ttcttgaata	tatctggact	ggtaaaatct	tgatgaggct											1455
	cataaaatga	gtttgggaat	tgtgtatagc	tgattttttg	tgggaaactg	tttacttcat											1515
	tcaaagggttc	ttgagactct	tgatatttct	gtcttctcct	tgtgctttcc	tatggaaaaa											1575
	atacatatat	agtttagttt	gttagacgtg	agttatccaa	gtattttattt	tgtgtagtgt											1635
	gtaagaatgc	taaataaaat	gttatacagg	aaaaaaaaaa	aaaaatgcga	aa											1687
45																	
	<210>	148															
	<211>	1747															
	<212>	DNA															
	<213>	Homo sapiens															
50																	
	<220>																
	<221>	CDS															
	<222>	70..1596															
55	<400>	148															
	gttggggcggc	cggtagctgt	tgctgttggg	ggacccctc	attcctgccg	ctgccgtccc											60
	tgctgcctc	atg	gcg	gcc	atc	gga	gtt	cac	ctg	ggc	tgc	acc	tca	gcc	tgt		111
		Met	Ala	Ala	Ile	Gly	Val	His	Leu	Gly	Cys	Thr	Ser	Ala	Cys		
		1				5					10						
60	gtg	gcc	gtc	tat	aag	gat	ggc	cgg	gct	ggg	gtg	gtt	gca	aat	gat	gcc	159
	Val	Ala	Val	Tyr	Lys	Asp	Gly	Arg	Ala	Gly	Val	Val	Ala	Asn	Asp	Ala	
		15				20					25				30		
	ggg	gac	cga	gtt	act	cca	gct	gtt	gtt	gct	tac	tca	gaa	aat	gaa	gag	207

	Gly	Asp	Arg	Val	Thr	Pro	Ala	Val	Val	Ala	Tyr	Ser	Glu	Asn	Glu	Glu	
					35					40					45		
	att	ggt	gga	ttg	gca	gca	aaa	caa	agt	aga	ata	aga	aat	att	tca	aat	255
5	Ile	Val	Gly	Leu	Ala	Ala	Lys	Gln	Ser	Arg	Ile	Arg	Asn	Ile	Ser	Asn	
				50				55					60				
	aca	gta	atg	aaa	gta	aag	cag	atc	ctg	ggc	aga	agc	tcc	agt	gat	cca	303
	Thr	Val	Met	Lys	Val	Lys	Gln	Ile	Leu	Gly	Arg	Ser	Ser	Ser	Asp	Pro	
				65				70					75				
	caa	gct	cag	aaa	tac	atc	gcg	gaa	agt	aaa	tgt	tta	gtc	att	gaa	aaa	351
10	Gln	Ala	Gln	Lys	Tyr	Ile	Ala	Glu	Ser	Lys	Cys	Leu	Val	Ile	Glu	Lys	
		80					85					90					
	aat	ggg	aaa	tta	cga	tat	gaa	ata	gat	act	gga	gaa	gaa	aca	aaa	ttt	399
	Asn	Gly	Lys	Leu	Arg	Tyr	Glu	Ile	Asp	Thr	Gly	Glu	Glu	Thr	Lys	Phe	
		95				100					105					110	
15	ggt	aac	cca	gaa	gat	ggt	gcc	aga	ctg	ata	ttt	agt	aaa	atg	aaa	gaa	447
	Val	Asn	Pro	Glu	Asp	Val	Ala	Arg	Leu	Ile	Phe	Ser	Lys	Met	Lys	Glu	
					115						120					125	
	acg	gca	cat	tct	gta	ttg	ggc	tca	gat	gca	aat	gat	gta	ggt	att	act	495
20	Thr	Ala	His	Ser	Val	Leu	Gly	Ser	Asp	Ala	Asn	Asp	Val	Val	Ile	Thr	
				130					135					140			
	gtc	ccg	ttt	gat	ttt	gga	gaa	aag	caa	aaa	aat	gct	ctt	gga	gaa	gca	543
	Val	Pro	Phe	Asp	Phe	Gly	Glu	Lys	Gln	Lys	Asn	Ala	Leu	Gly	Glu	Ala	
				145				150					155				
	gct	aga	gct	gct	gga	ttt	aat	ggt	ttg	cga	tta	att	cac	gaa	ccg	tct	591
25	Ala	Arg	Ala	Ala	Gly	Phe	Asn	Val	Leu	Arg	Leu	Ile	His	Glu	Pro	Ser	
		160					165					170					
	gca	gct	ctt	ctt	gct	tat	gga	att	gga	caa	gac	tcc	cct	act	gga	aaa	639
	Ala	Ala	Leu	Leu	Ala	Tyr	Gly	Ile	Gly	Gln	Asp	Ser	Pro	Thr	Gly	Lys	
		175				180					185					190	
30	agc	aat	att	ttg	gtg	ttt	aag	ctt	gga	gga	aca	tcc	tta	tct	ctc	agc	687
	Ser	Asn	Ile	Leu	Val	Phe	Lys	Leu	Gly	Gly	Thr	Ser	Leu	Ser	Leu	Ser	
					195					200						205	
	gtc	atg	gaa	ggt	aac	agt	gga	ata	tat	cgg	ggt	ctt	tca	aca	aac	act	735
	Val	Met	Glu	Val	Asn	Ser	Gly	Ile	Tyr	Arg	Val	Leu	Ser	Thr	Asn	Thr	
35				210					215							220	
	gat	gat	aac	atc	ggg	ggg	gca	cat	ttc	aca	gaa	acc	tta	gca	cag	tat	783
	Asp	Asp	Asn	Ile	Gly	Gly	Ala	His	Phe	Thr	Glu	Thr	Leu	Ala	Gln	Tyr	
				225				230					235				
	cta	gct	tct	gag	ttc	caa	aga	tcc	ttc	aaa	cat	gat	gtg	aga	gga	aat	831
40	Leu	Ala	Ser	Glu	Phe	Gln	Arg	Ser	Phe	Lys	His	Asp	Val	Arg	Gly	Asn	
		240					245					250					
	gcg	cga	gcc	atg	atg	aaa	tta	acg	aac	agt	gct	gaa	gta	gcg	aaa	cat	879
	Ala	Arg	Ala	Met	Met	Lys	Leu	Thr	Asn	Ser	Ala	Glu	Val	Ala	Lys	His	
		255				260					265					270	
45	tct	ttg	tca	acc	ttg	gga	agt	gcc	aac	tgt	ttt	ctt	gac	tca	tta	tat	927
	Ser	Leu	Ser	Thr	Leu	Gly	Ser	Ala	Asn	Cys	Phe	Leu	Asp	Ser	Leu	Tyr	
					275					280						285	
	gaa	ggg	caa	gat	ttt	gat	tgc	aat	gtg	tcc	aga	gca	aga	ttt	gaa	ctt	975
	Glu	Gly	Gln	Asp	Phe	Asp	Cys	Asn	Val	Ser	Arg	Ala	Arg	Phe	Glu	Leu	
50				290					295					300			
	ctt	tgt	tct	cca	ctt	ttt	aat	aag	tgt	ata	gaa	gca	atc	aga	gga	ctc	1023
	Leu	Cys	Ser	Pro	Leu	Phe	Asn	Lys	Cys	Ile	Glu	Ala	Ile	Arg	Gly	Leu	
				305				310					315				
	tta	gat	caa	aat	gga	ttt	aca	aca	gat	gat	atc	aac	aag	gtt	gtc	ctt	1071
55	Leu	Asp	Gln	Asn	Gly	Phe	Thr	Thr	Asp	Asp	Ile	Asn	Lys	Val	Val	Leu	
		320					325					330					
	tgt	gga	ggg	tct	tct	cga	atc	cca	aag	cta	cag	caa	ctg	att	aaa	gat	1119
	Cys	Gly	Gly	Ser	Ser	Arg	Ile	Pro	Lys	Leu	Gln	Gln	Leu	Ile	Lys	Asp	
		335				340					345					350	
60	ctt	ttc	cca	gct	ggt	gag	ctt	ctc	aat	tct	atc	cct	cct	gat	gaa	gtg	1167
	Leu	Phe	Pro	Ala	Val	Glu	Leu	Leu	Asn	Ser	Ile	Pro	Pro	Asp	Glu	Val	
					355					360					365		
	atc	cct	att	ggg	gca	gct	ata	gaa	gca	gga	att	ctt	att	ggg	aaa	gaa	1215

	Ile	Pro	Ile	Gly	Ala	Ala	Ile	Glu	Ala	Gly	Ile	Leu	Ile	Gly	Lys	Glu	
				370					375					380			
	aac	ctg	ttg	gtg	gaa	gac	tct	ctt	atg	ata	gag	tgt	tca	gcc	aga	gat	1263
5	Asn	Leu	Leu	Val	Glu	Asp	Ser	Leu	Met	Ile	Glu	Cys	Ser	Ala	Arg	Asp	
			385					390					395				
	att	tta	gtt	aag	ggt	gtg	gac	gaa	tca	gga	gcc	agt	aga	ttc	aca	gtg	1311
	Ile	Leu	Val	Lys	Gly	Val	Asp	Glu	Ser	Gly	Ala	Ser	Arg	Phe	Thr	Val	
			400				405					410					
	ctg	ttt	cca	tca	ggg	act	cct	ttg	cca	gct	cga	aga	caa	cac	aca	ttg	1359
10	Leu	Phe	Pro	Ser	Gly	Thr	Pro	Leu	Pro	Ala	Arg	Arg	Gln	His	Thr	Leu	
						420					425				430		
	caa	gcc	cct	gga	agc	ata	tct	tca	gtg	tgc	ctt	gaa	ctc	tat	gag	tct	1407
	Gln	Ala	Pro	Gly	Ser	Ile	Ser	Ser	Val	Cys	Leu	Glu	Leu	Tyr	Glu	Ser	
					435					440					445		
15	gat	ggg	aag	aac	tct	gcc	aaa	gag	gaa	acc	aag	ttt	gca	cag	gtt	gta	1455
	Asp	Gly	Lys	Asn	Ser	Ala	Lys	Glu	Glu	Thr	Lys	Phe	Ala	Gln	Val	Val	
				450				455					460				
	ctc	cag	gat	tta	gat	aaa	aaa	gaa	aat	gga	tta	cgt	gat	ata	tta	gct	1503
20	Leu	Gln	Asp	Leu	Asp	Lys	Lys	Glu	Asn	Gly	Leu	Arg	Asp	Ile	Leu	Ala	
			465					470					475				
	gtt	ctt	act	atg	aaa	agg	gat	gga	tct	tta	cat	gtg	aca	tgc	aca	gat	1551
	Val	Leu	Thr	Met	Lys	Arg	Asp	Gly	Ser	Leu	His	Val	Thr	Cys	Thr	Asp	
			480				485					490					
	caa	gaa	act	gga	aaa	tgt	gaa	gca	atc	tct	att	gag	ata	gca	tct		1596
25	Gln	Glu	Thr	Gly	Lys	Cys	Glu	Ala	Ile	Ser	Ile	Glu	Ile	Ala	Ser		
						500					505						
	tagtg	ttt	tta	gagaa	atcaa	gaatt	ttt	tta	aaaca	agaat	atcaac	attt	gg	ttt	gtgt		1656
	ata	agt	gg	t	g	t	t	t	a	a	t	a	a	a	a	a	1716
	act	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	1747
30																	
	<210> 149																
	<211> 658																
	<212> DNA																
	<213> Homo sapiens																
35																	
	<220>																
	<221> CDS																
	<222> 129..362																
40	<400> 149																
	agtcagg	aaa	atgaag	ctga	acatca	agtc	ccagca	agaa	aagaagg	aaa	aggaat	gcca					60
	agcag	tc	atg	ttt	gaagg	ggtc	gagca	agatg	gaatcagg	aa	taagg	agtca					120
	gtggg	acc	atg	tac	aac	act	gga	aga	cac	gta	tcc	ctt	cgc	ctg	gac	aag	170
			Met	Tyr	Asn	Thr	Gly	Arg	His	Val	Ser	Leu	Arg	Leu	Asp	Lys	
45			1			5					10						
	gag	cac	ttg	gtc	aac	ata	tct	gga	ggg	ccc	atg	aca	tac	agc	cac	cgg	218
	Glu	His	Leu	Val	Asn	Ile	Ser	Gly	Gly	Pro	Met	Thr	Tyr	Ser	His	Arg	
			15			20					25					30	
	ctg	gag	gag	atc	cga	cta	cac	ttt	ggg	agt	gag	gac	agc	caa	ggg	tcg	266
50	Leu	Glu	Glu	Ile	Arg	Leu	His	Phe	Gly	Ser	Glu	Asp	Ser	Gln	Gly	Ser	
					35					40					45		
	gag	cac	ctc	ctc	aat	gga	cag	gcc	ttc	tct	ggg	gag	ctt	caa	gag	agg	314
	Glu	His	Leu	Leu	Asn	Gly	Gln	Ala	Phe	Ser	Gly	Glu	Leu	Gln	Glu	Arg	
					50					55					60		
55	gat	ttg	ttc	atc	ttg	ttg	act	tct	gta	tca	gga	cat	ctg	ccc	gat	aca	362
	Asp	Leu	Phe	Ile	Leu	Leu	Thr	Ser	Val	Ser	Gly	His	Leu	Pro	Asp	Thr	
			65					70									
	tag	aaaa	agt	ctg	ctg	accc	ctga	attaca	gtatg	agcca	ttc	gga	aatgc	atttct	ctttt		422
	aaa	agt	ttctc	gcctc	attca	gtgtc	tggaa	cacagt	gggt	gctcccc	aat	aggtg	acacc				482
60	ttc	ctca	agt	ttc	ctt	ggga	gaacag	actc	aatgtc	ggat	ccacaa	agga	gacctg	caca			542
	tac	cta	accc	ctatt	ttctg	c	agaag	ctgaa	ggctg	tatta	tctatt	gctt	gcata	ataaaa			602
	tattg	cataa	cgaca	acaat	agta	aaaaaaa	aaaaaaaa	aaaaaaaa	gaaaa	aaaaa							658

<210> 150
 <211> 2045
 <212> DNA
 <213> Homo sapiens
 5
 <220>
 <221> CDS
 <222> 109..594
 10 <400> 150
 attattacta caggaaaaaac tgttctcttc tgtggcacag agaaccctgc ttcaaagcag 60
 aagtagcagt tccggagtc agctggctaa aactcatccc agaggata atg gca acc 117
 Met Ala Thr
 1
 15 cat gcc tta gaa atc gct ggg ctg ttt ctt ggt ggt gtt gga atg gtg 165
 His Ala Leu Glu Ile Ala Gly Leu Phe Leu Gly Gly Val Gly Met Val
 5 10 15
 ggc aca gtg gct gtc act gtc atg cct cag tgg ata gtg tcg gcc ttc 213
 Gly Thr Val Ala Val Thr Val Met Pro Gln Trp Ile Val Ser Ala Phe
 20 25 30 35
 att gaa aac aac atc gtg gtt ttt gaa aac ttc tgg gaa gga ctg tgg 261
 Ile Glu Asn Asn Ile Val Val Phe Glu Asn Phe Trp Glu Gly Leu Trp
 40 45 50
 atg aat tgc gtg agg cag gct aac atc agg atg cag tgc aaa atc tat 309
 Met Asn Cys Val Arg Gln Ala Asn Ile Arg Met Gln Cys Lys Ile Tyr
 55 60 65
 gat tcc ctg ctg gct ctt tct ccg gac cta cag gca gcc aga gga ctg 357
 Asp Ser Leu Leu Ala Leu Ser Pro Asp Leu Gln Ala Ala Arg Gly Leu
 70 75 80
 30 atg tgt gct gct tcc gtg atg tcc ttc ttg gct ttc atg atg gcc atc 405
 Met Cys Ala Ala Ser Val Met Ser Phe Leu Ala Phe Met Met Ala Ile
 85 90 95
 ctt ggc atg aaa tgc acc agg tgc acg ggg gac aat gag aag gtg aag 453
 Leu Gly Met Lys Cys Thr Arg Cys Thr Gly Asp Asn Glu Lys Val Lys
 100 105 110 115
 35 gct cac att ctg ctg acg gct gga atc atc ttc atc atc acg ggc atg 501
 Ala His Ile Leu Leu Thr Ala Gly Ile Ile Phe Ile Ile Thr Gly Met
 120 125 130
 gtg gtg ctc atc cct gtg agc tgg gtt gcc aat gcc atc atc aga gat 549
 Val Val Leu Ile Pro Val Ser Trp Val Ala Asn Ala Ile Ile Arg Asp
 135 140 145
 ttc tat aac cca ata gtg aat gtt gcc caa aaa cgt gag ctt gga 594
 Phe Tyr Asn Pro Ile Val Asn Val Ala Gln Lys Arg Glu Leu Gly
 150 155 160
 45 taagctctct acttaggatg gaccacggca ctggtgctga ttggtggagg agctctgttc 654
 tgcgtcggtt tttgttgcaa cgaaaagagc agtagctaca gatactcgat accttcccat 714
 cgcacaaccc aaaaaagtta tcacaccgga aagaagtcac cgagcgtcta ctccagaagt 774
 cagtatgtgt agttgtgtat gtttttttaa ctttactata aagccatgca aatgacaaaa 834
 atctatatata ctttctcaaa atggacccca aagaaacttt gatttactgt tcttaactgc 894
 50 ctaatcttaa ttacaggaac tgtgcatcag ctatttatga ttctataagc tatttcagca 954
 gaatgagata ttaaataccaa tgctttgatt gttctagaaa gtatagtaat ttgttttcta 1014
 aggtggttca agcatctact ctttttatca tttacttcaa aatgacattg ctaaagactg 1074
 cattatttta ctactgtaat ttctccacga catagcatta tgtacataga tgagtgtaac 1134
 atttatatct cacatagaga catgcttata tgggtttatt taaaatgaaa tgccagtcca 1194
 55 ttacactgaa taaatagaac tcaactattg cttttcaggg aaatcatgga taggggtgaa 1254
 gaaggttact attaatgtt taaaaacagc ttagggatta atgtcctcca tttataatga 1314
 agattaaaat gaaggcttta atcagcattg taaaggaaat tgaatggctt tctgatatgc 1374
 tgttttttag cctaggagtt agaaatccta acttctttat cctcttctcc cagaggcttt 1434
 ttttttcttg tgtattaaat taacattttt aaaaagcaga tattttgtca aggggctttg 1494
 60 cattcaaact gcttttccag ggctatactc agaagaaaga taaaagtgtg atctaagaaa 1554
 aagtgatggt tttaggaaag tgaaaatatt tttgtttttg tatttgaaga agaagatgac 1614
 attttgacaa gaaatcatat atgtatggat atattttta aagtatttga gtacagactt 1674
 tgagggtttca tcaatataaa taaaagagca gaaaaatatg tcttggtttt catttgctta 1734

	ccaaaaaaac	aacaacaaaa	aaagttgtcc	tttgagaact	tcacctgctc	ctatgtgggt	1794	
	acctgagtca	aaattgtcat	ttttgttctg	tgaaaaataa	atttccttct	tgtaccattt	1854	
	ctgttttagtt	ttactaaaat	ctgtaaatac	tgtatttttc	tgtttattcc	aaatttgatg	1914	
	aaactgacaa	tccaatttga	aagtttgtgt	cgacgtctgt	ctagcttaaa	tgaatgtgtt	1974	
5	ctatttgctt	tatacattta	tattaataaa	ttgtacattt	ttctaattat	ttggaaaaaa	2034	
	aaaaaaaaaa	a					2045	
	<210> 151							
	<211> 788							
10	<212> DNA							
	<213> Homo sapiens							
	<220>							
	<221> CDS							
15	<222> 150..587							
	<400> 151							
	attttcaa	at	ttacccctct	gtgacttgta	agccatgcaa	ttcgtagggc	taaatatgca	60
	gttggttcgat	ttcacgggtt	ggaatctctt	gtcaaggac	tgaggactctt	caattaatct		120
20	gacatttcac	aaatccaaaa	ttgccgtgg	atg aac tct	tta ctt cac ttc	ggg		173
				Met Asn Ser	Leu Leu His Phe Gly			
				1	5			
	ata ttg ctg gag ctg agt ctc ctg	aaa cag ttt aag tct gta tat gtt						221
	Ile Leu Leu Glu Leu Ser Leu Leu	Lys Gln Phe Lys Ser Val Tyr Val						
25	10		15		20			
	cct gga aat cat acc cac cag gca	tct tat aag cca ttg ttg aag caa						269
	Pro Gly Asn His Thr His Gln Ala	Ser Tyr Lys Pro Leu Leu Lys Gln						
	25		30		35		40	
	gtt gtg gag gaa ata ttt cat ccc	gag agg cca gat tcc gtt gat att						317
30	Val Val Glu Glu Ile Phe His Pro	Glu Arg Pro Asp Ser Val Asp Ile						
			45		50		55	
	gaa cac atg tct tca ggc ctc act	gat ctc ctt aaa act gga ttt agc						365
	Glu His Met Ser Ser Gly Leu Thr	Asp Leu Leu Lys Thr Gly Phe Ser						
			60		65		70	
35	atg ttc atg aag gtg agc cgg cct	cat cct agt gac tac ccc ctc ctg						413
	Met Phe Met Lys Val Ser Arg Pro	His Pro Ser Asp Tyr Pro Leu Leu						
			75		80		85	
	atc ctc ttt gtg gta ggt ggg gtc	aca gtc tct gaa gtg aaa atg gtc						461
	Ile Leu Phe Val Val Gly Gly Val	Thr Val Ser Glu Val Lys Met Val						
40	90		95		100			
	aaa gat ctt gtg gca tcg ttg aag	cca gga acc cag gta atc gtg ctg						509
	Lys Asp Leu Val Ala Ser Leu Lys	Pro Gly Thr Gln Val Ile Val Leu						
			105		110		115	120
	tcc aca cga ctc ctg aag cca ctt	aac att cct gag ctg tta ttt gca						557
45	Ser Thr Arg Leu Leu Lys Pro Leu	Asn Ile Pro Glu Leu Leu Phe Ala						
			125		130		135	
	act gac cga ctg cat cca gac ctt	ggc ttc tgagcatccg ctaagaagat						607
	Thr Asp Arg Leu His Pro Asp Leu	Gly Phe						
			140		145			
50	aagacctact	caagctggaa	atgccgatgc	aatttttctgc	caccactcca	aatactcctc		667
	cacaaccagc	gtccctgtca	ctaattgcga	gaatgatgga	attctgcctg	aagggctctg		727
	atacctactc	agtgaggtac	stttgcttgg	attgctgtga	ttccaaaaaa	aaaaaaaaaa		787
	a							788
	<210> 152							
	<211> 1931							
	<212> DNA							
	<213> Homo sapiens							
	<220>							
	<221> CDS							
60	<222> 173..847							

<400> 152													
	actggatact	atctggccag	aagtagcaaa	gcagctctta	tttgaaaaac	cactgggttc							60
	cgagttcatt	actacaggaa	aaactgttct	cttctgtggc	acagagaacc	ctgcttcaaa							120
	gcagaagtag	cagttccgga	gtccagctgg	ctaaaactca	tcccagagga	ta atg gca							178
5						Met Ala							
						1							
	acc cat gcc tta gaa atc gct ggg ctg ttt ctt ggt ggt gtt gga atg												226
	Thr His Ala Leu Glu Ile Ala Gly Leu Phe Leu Gly Gly Val Gly Met												
		5				10				15			
10	gtg ggc aca gtg gct gtc act gtc atg cct cag tgg aga gtg tcg gcc												274
	Val Gly Thr Val Ala Val Thr Val Met Pro Gln Trp Arg Val Ser Ala												
		20				25				30			
	ttc att gaa aac aac atc gtg gtt ttt gaa aac ttc tgg gaa gga ctg												322
	Phe Ile Glu Asn Asn Ile Val Val Phe Glu Asn Phe Trp Glu Gly Leu												
15		35				40				45			50
	tgg atg aat tgc gtg agg cag gct aac atc agg atg cag tgc aaa atc												370
	Trp Met Asn Cys Val Arg Gln Ala Asn Ile Arg Met Gln Cys Lys Ile												
						55				60			65
	tat gat tcc ctg gct ctt tct ccg gac cta cag gca gcc aga gga												418
20	Tyr Asp Ser Leu Leu Ala Leu Ser Pro Asp Leu Gln Ala Ala Arg Gly												
						70				75			80
	ctg atg tgt gct gct tcc gtg atg tcc ttc ttg gct ttc atg atg gcc												466
	Leu Met Cys Ala Ala Ser Val Met Ser Phe Leu Ala Phe Met Met Ala												
						85				90			95
25	atc ctt ggc atg aaa tgc acc agg tgc acg ggg gac aat gag aag gtg												514
	Ile Leu Gly Met Lys Cys Thr Arg Cys Thr Gly Asp Asn Glu Lys Val												
						100				105			110
	aag gct cac att ctg ctg acg gct gga atc atc ttc atc atc gcg ggc												562
	Lys Ala His Ile Leu Leu Thr Ala Gly Ile Phe Ile Ile Ala Gly												
30						115				120			125
	atg gtg gtg ctc atc cct gtg agc tgg gtt gcc aat gcc atc atc aga												610
	Met Val Val Leu Ile Pro Val Ser Trp Val Ala Asn Ala Ile Ile Arg												
						135				140			145
	gat ttc tat aac cca ata gtg aat gtt gcc caa aaa cgt gag ctt gga												658
35	Asp Phe Tyr Asn Pro Ile Val Asn Val Ala Gln Lys Arg Glu Leu Gly												
						150				155			160
	gaa gct ctc tac tta gga tgg acc acg gca ctg gtg ctg att gtt gga												706
	Glu Ala Leu Tyr Leu Gly Trp Thr Thr Ala Leu Val Leu Ile Val Gly												
						165				170			175
40	gga gct ctg ttc tgc tgc gtt ttt tgt tgc aac gaa aag agc agt agc												754
	Gly Ala Leu Phe Cys Cys Val Phe Cys Cys Asn Glu Lys Ser Ser Ser												
						180				185			190
	tac aga tac tcg ata cct tct cat cgc aca acc caa aaa agt tat cac												802
	Tyr Arg Tyr Ser Ile Pro Ser His Arg Thr Thr Gln Lys Ser Tyr His												
45						195				200			205
	acc gga aag aag tca ccg agc gtc tac tcc aga agt cag tat gtg												847
	Thr Gly Lys Lys Ser Pro Ser Val Tyr Ser Arg Ser Gln Tyr Val												
						215				220			225
	tagttgtgta tgttttttta actttactat aaagctatgc aaatgacaaa aatctatatt												907
50	acttttctcaa aatggacccc aaagaaactt tgatttactg ttcttaactg cctaattctta												967
	attacaggaa ctgtgcatca gctatttatg attctataag ctatttcagc agaatgagat												1027
	attaaatcca atgctttgat tgttctagaa agtatagtaa tttgttttct aagggtggttc												1087
	aagcatctac tctttttatc atttacttca aaatgacatt gctaaagact gcattatttt												1147
	actactgtaa tttctccatg acatagcatt atgtacatag atgagtgtaa catttatatc												1207
55	tcacatagag acatgcttat atggttttat ttaaaatgaa atgccagtcc attacactga												1267
	tataaatagaa ctcaactatt gcttttcagg gaaatcatgg ataggggtga agaagggttac												1327
	tattaattgt ttaaaaaacag cttagggatt aatgtctctc atttataatg aagattaaaa												1387
	tgaaggcttt aatcagcatt gtaaaaggaaa ttgaatggct ttctgatatg ctgtttttta												1447
	gcctaggagt tagaaatcct aacttcttta tcctcttctc ccagaggctt tttttttctt												1507
60	gtgtattaaa ttaacatttt taaaaagcag atattttgtc aaggggcttt gcattcaaac												1567
	tgctttttcca gggctatact cagaagaaag ataaaaagtg gatctaagaa aaagtgatgg												1627
	ttttaggaaa gtgaaaatat ttttgttttt gtatttgaag aagaatgatg cattttgaca												1687
	agaaatcata tatgtatgga tatattttta taagtatttg agtacagact ttgaggtttc												1747

atcaatataa ataaaagagc agaaaaatat gtcttggttt tcatttgctt accaaaaaaaa 1807
 caacaacaaa aaaagtgtc ctttgagaac ttcacctgct cctatgtggg tacctgagtc 1867
 aaaattgtca tttttgttct gtgaaaaata aatttccttc ttgtaccaa aaaaaaaaaa 1927
 aaaa 1931

5

<210> 153
 <211> 514
 <212> DNA
 <213> Homo sapiens

10

<220>
 <221> CDS
 <222> 100..441

15

<400> 153
 ataccaggca ctttagaacc agagactctg ctgcttttcc tgggcagggc ctgcttgctc 60
 cagctctcaa gtctgacttg catctacact gcgggcaag atg cgg ctg caa gac 114
 Met Arg Leu Gln Asp
 1 5

20

cgc atc gcc acg ttc ttc ttc cca aaa ggc atg atg ctc acc acg gct 162
 Arg Ile Ala Thr Phe Phe Phe Pro Lys Gly Met Met Leu Thr Thr Ala
 10 15 20
 gcg ctg atg ctc ttc ttc tta cac ctg ggc atc ttc atc aga gac gtg 210
 Ala Leu Met Leu Phe Phe Leu His Leu Gly Ile Phe Ile Arg Asp Val
 25 25 30 35

25

cac aac ttc tgc atc acc tac cac tat gac cac atg agc ttt cac tac 258
 His Asn Phe Cys Ile Thr Tyr His Tyr Asp His Met Ser Phe His Tyr
 40 45 50

30

acg gtc gtc ctg atg ttc tcc cag gtg atc agc atc tgc tgg gct gcc 306
 Thr Val Val Leu Met Phe Ser Gln Val Ile Ser Ile Cys Trp Ala Ala
 55 60 65

35

atg ggg tca ctc tat gct gag atg aca gaa aac aat gct caa cgg agc 354
 Met Gly Ser Leu Tyr Ala Glu Met Thr Glu Asn Asn Ala Gln Arg Ser
 70 75 80 85

40

cat gtt ctt caa ccg cct gtc ctt gga gtt tct ggc cat cga gta ccg 402
 His Val Leu Gln Pro Pro Val Leu Gly Val Ser Gly His Arg Val Pro
 90 95 100

45

gga gga gca cca ctg agg cct ggg gag tgc gaa cag ggc taaggagggg 451
 Gly Gly Ala Pro Leu Arg Pro Gly Glu Ser Glu Gln Gly
 105 110

50

gaagcaaaag gctgcctcgg gtgttttaaat aaagttgttg tttattccaa aaaaaaaaaa 511
 aaa 514

<210> 154
 <211> 1183
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 32..1132

55

<400> 154
 acttctttcc tgcctctgat tccgggctgt c atg gcg acc ccc aac aat ctg 52
 Met Ala Thr Pro Asn Asn Leu
 1 5

60

acc ccc acc aac tgc agc tgg tgg ccc atc tcc gcg ctg gag agc gat 100
 Thr Pro Thr Asn Cys Ser Trp Trp Pro Ile Ser Ala Leu Glu Ser Asp
 10 15 20

gcg gcc aag cca gcg gag gcc ccc gac gct ccc gag gcg gcc agc ccc 148
 Ala Ala Lys Pro Ala Glu Ala Pro Asp Ala Pro Glu Ala Ala Ser Pro
 25 30 35

gcc cat tgg ccc agg gag agc ctg gtt ctg tac cac tgg acc cag tcc 196

	Ala	His	Trp	Pro	Arg	Glu	Ser	Leu	Val	Leu	Tyr	His	Trp	Thr	Gln	Ser	
	40					45					50					55	
	ttc	agc	tcg	cag	aag	gtg	cgg	ctg	gtg	atc	gcc	gag	aag	ggc	ctg	gtg	244
5	Phe	Ser	Ser	Gln	Lys	Val	Arg	Leu	Val	Ile	Ala	Glu	Lys	Gly	Leu	Val	
					60					65					70		
	tgc	gag	gag	cgg	gac	gtg	agc	ctg	cca	cag	agc	gag	cac	aag	gag	ccc	292
	Cys	Glu	Glu	Arg	Asp	Val	Ser	Leu	Pro	Gln	Ser	Glu	His	Lys	Glu	Pro	
					75				80					85			
	tgg	ttc	atg	cgg	ctc	aac	ctg	ggc	gag	gag	gtg	ccc	gtc	atc	atc	cac	340
10	Trp	Phe	Met	Arg	Leu	Asn	Leu	Gly	Glu	Glu	Val	Pro	Val	Ile	Ile	His	
			90					95					100				
	cgc	gac	aac	atc	atc	agt	gac	tat	gac	cag	atc	att	gac	tat	gtg	gag	388
	Arg	Asp	Asn	Ile	Ile	Ser	Asp	Tyr	Asp	Gln	Ile	Ile	Asp	Tyr	Val	Glu	
		105					110					115					
15	cgc	acc	ttc	aca	gga	gag	cac	gtg	gtg	gcc	ctg	atg	ccc	gag	gtg	ggc	436
	Arg	Thr	Phe	Thr	Gly	Glu	His	Val	Val	Ala	Leu	Met	Pro	Glu	Val	Gly	
						125				130						135	
	agc	ctg	cag	cac	gca	cgg	gtg	ctg	cag	tac	cgg	gag	ctg	ctg	gac	gca	484
	Ser	Leu	Gln	His	Ala	Arg	Val	Leu	Gln	Tyr	Arg	Glu	Leu	Leu	Asp	Ala	
20					140					145					150		
	ctg	ccc	atg	gat	gcc	tac	acg	cat	ggc	tgc	atc	ctg	cat	ccc	gag	ctc	532
	Leu	Pro	Met	Asp	Ala	Tyr	Thr	His	Gly	Cys	Ile	Leu	His	Pro	Glu	Leu	
					155				160					165			
	acc	acc	gac	tcc	atg	atc	ccc	aag	tac	gcc	acg	gcc	gag	atc	cgc	aga	580
25	Thr	Thr	Asp	Ser	Met	Ile	Pro	Lys	Tyr	Ala	Thr	Ala	Glu	Ile	Arg	Arg	
			170					175					180				
	cat	tta	gcc	aac	gcc	acc	acg	gac	ctc	atg	aaa	ctg	gac	cat	gaa	gag	628
	His	Leu	Ala	Asn	Ala	Thr	Thr	Asp	Leu	Met	Lys	Leu	Asp	His	Glu	Glu	
		185					190					195					
30	gag	ccc	cag	ctc	tcc	gag	ccc	tac	ctt	tct	aaa	caa	aag	aag	ctc	atg	676
	Glu	Pro	Gln	Leu	Ser	Glu	Pro	Tyr	Leu	Ser	Lys	Gln	Lys	Lys	Leu	Met	
						205					210					215	
	gtc	aag	atc	ttg	gag	cat	gat	gat	gtg	agc	tac	ctg	aag	aag	atc	ctc	724
	Val	Lys	Ile	Leu	Glu	His	Asp	Asp	Val	Ser	Tyr	Leu	Lys	Lys	Ile	Leu	
35					220					225					230		
	ggg	gaa	ctg	gcc	atg	gtg	ctg	gac	cag	att	gag	gcg	gag	ctg	gag	aag	772
	Gly	Glu	Leu	Ala	Met	Val	Leu	Asp	Gln	Ile	Glu	Ala	Glu	Leu	Glu	Lys	
					235				240					245			
	agg	aag	ctg	gag	aac	gag	ggg	cag	aaa	tgc	gag	ctg	tgg	ctc	tgt	ggc	820
40	Arg	Lys	Leu	Glu	Asn	Glu	Gly	Gln	Lys	Cys	Glu	Leu	Trp	Leu	Cys	Gly	
			250					255					260				
	tgt	gcc	ttc	acc	ctc	gct	gat	gtc	ctc	ctg	gga	gcc	acc	ctg	cac	cgc	868
	Cys	Ala	Phe	Thr	Leu	Ala	Asp	Val	Leu	Leu	Gly	Ala	Thr	Leu	His	Arg	
		265					270					275					
45	ctc	aag	ttc	ctg	gga	ctg	tcc	aag	aaa	tac	tgg	gaa	gat	ggc	agc	cgg	916
	Leu	Lys	Phe	Leu	Gly	Leu	Ser	Lys	Lys	Tyr	Trp	Glu	Asp	Gly	Ser	Arg	
						285					290					295	
	ccc	aac	ctg	cag	tcc	ttc	ttt	gag	agg	gtc	cag	aga	cgc	ttt	gcc	ttc	964
	Pro	Asn	Leu	Gln	Ser	Phe	Phe	Glu	Arg	Val	Gln	Arg	Arg	Phe	Ala	Phe	
50					300					305					310		
	cgg	aaa	gtc	ctg	ggt	gac	atc	cac	acc	acc	ctg	ctg	tcg	gcc	gtc	atc	1012
	Arg	Lys	Val	Leu	Gly	Asp	Ile	His	Thr	Thr	Leu	Leu	Ser	Ala	Val	Ile	
					315				320					325			
	ccc	aac	gct	ttc	cgg	ctg	gtc	aag	agg	aaa	ccc	cca	tcc	ttc	ttc	ggg	1060
55	Pro	Asn	Ala	Phe	Arg	Leu	Val	Lys	Arg	Lys	Pro	Pro	Ser	Phe	Phe	Gly	
			330					335					340				
	gcg	tcc	ttc	ctc	atg	ggc	tcc	ctg	ggt	ggg	atg	ggc	tac	ttt	gcc	tac	1108
	Ala	Ser	Phe	Leu	Met	Gly	Ser	Leu	Gly	Gly	Met	Gly	Tyr	Phe	Ala	Tyr	
		345					350					355					
60	tgg	tac	ctc	aag	aaa	aaa	tac	atc	tagggccagg	cctggggcctt	ggtgtctgac						1162
	Trp	Tyr	Leu	Lys	Lys	Lys	Tyr	Ile									
		360				365											
	tgtcaaaaaa	aaaaaaaaaa	a														1183

<210> 155
 <211> 1545
 <212> DNA
 5 <213> Homo sapiens

<220>
 <221> CDS
 <222> 160..996

10 <400> 155
 acacagcatg catttcttca acaagcgact cagaaggcac ttgcacatcg ttgctgttct 60
 gcctcttttg ttcagcatga ttaccagag gcgcaccgt gccgtggcct gcccgtcgtc 120
 tatgcacccg tgctgtggcg tgcccgtcgt ctgtgtggc atg cct gtc tgt gca 174
 Met Pro Val Cys Ala
 1 5

15 ccc gtg ctg tgg cgt gcc cgt cgt ctg tgt ggc atg cct gtc tgt gca 222
 Pro Val Leu Trp Arg Ala Arg Arg Leu Cys Gly Met Pro Val Cys Ala
 10 15 20

20 ccc gtg ccg tgg cgt gcc cgt cgt ctg tgc acc cgt gct gtg gtg tgc 270
 Pro Val Pro Trp Arg Ala Arg Arg Leu Cys Thr Arg Ala Val Val Cys
 25 30 35

cct tcg tct gtt cct ttt att gcc ggg cag ggt tgc acc cac atg tgc 318
 Pro Ser Ser Val Pro Phe Ile Ala Gly Gln Gly Cys Thr His Met Cys
 40 45 50

25 aag cca gcg acg gac ccc agg ttc acc cgt tca ccg ctg gct gga ggc 366
 Lys Pro Ala Thr Asp Pro Arg Phe Thr Arg Ser Pro Leu Ala Gly Gly
 55 60 65

30 gtg atc ctg ggt gtg gcc ctg tgg ctc cgc cat gac ccg cag acc acc 414
 Val Ile Leu Gly Val Ala Leu Trp Leu Arg His Asp Pro Gln Thr Thr
 70 75 80 85

aac ctc ctg tat ctg gag ctg gga gac aag ccc gcg ccc aac acc ttc 462
 Asn Leu Leu Tyr Leu Glu Leu Gly Asp Lys Pro Ala Pro Asn Thr Phe
 90 95 100

35 tat gta ggc atc tac atc ctc atc gct gtg ggc gct gtc atg atg ttc 510
 Tyr Val Gly Ile Tyr Ile Leu Ile Ala Val Gly Ala Val Met Met Phe
 105 110 115

gtt ggc ttc ctg ggc tgc tac ggg gcc atc cag gaa tcc cag tgc ctg 558
 Val Gly Phe Leu Gly Cys Tyr Gly Ala Ile Gln Glu Ser Gln Cys Leu
 120 125 130

40 ctg ggg acg ttc ttc act tgc ctg gtc atc ctg ttt gcc tgt gag gtg 606
 Leu Gly Thr Phe Phe Thr Cys Leu Val Ile Leu Phe Ala Cys Glu Val
 135 140 145

gcc gcc ggc atc tgg ggc ttt gtc aac aag gac cag atc gcc aag gat 654
 Ala Ala Gly Ile Trp Gly Phe Val Asn Lys Asp Gln Ile Ala Lys Asp
 150 155 160 165

gtg aag cag ttc tat gac cag gcc cta cag cag gcc gtg gtg gat gat 702
 Val Lys Gln Phe Tyr Asp Gln Ala Leu Gln Gln Ala Val Val Asp Asp
 170 175 180

50 gac gcc aac aac gcc aag gct gtg gtg aag acc ttc cac gag acg ctt 750
 Asp Ala Asn Asn Ala Lys Ala Val Val Lys Thr Phe His Glu Thr Leu
 185 190 195

gac tgc tgt ggc tcc agc aca ctg act gct ttg acc acc tca gtg ctc 798
 Asp Cys Cys Gly Ser Ser Thr Leu Thr Ala Leu Thr Thr Ser Val Leu
 200 205 210

55 aag aac aat ttg tgt ccc tcg ggc agc aac atc atc agc aac ctc ttc 846
 Lys Asn Asn Leu Cys Pro Ser Gly Ser Asn Ile Ile Ser Asn Leu Phe
 215 220 225

aag gag gac tgc cac cag aag atc gat gac ctc ttc tcc ggg aag ctg 894
 Lys Glu Asp Cys His Gln Lys Ile Asp Asp Leu Phe Ser Gly Lys Leu
 230 235 240 245

60 tac ctc atc ggc att gct gcc atc gtg gtc gct gtg atc atg atc ttc 942
 Tyr Leu Ile Gly Ile Ala Ala Ile Val Val Ala Val Ile Met Ile Phe

250 255 260
 gag atg atc ctg agc atg gtg ctg tgc tgt ggc atc cgg aac agc tcc 990
 Glu Met Ile Leu Ser Met Val Leu Cys Cys Gly Ile Arg Asn Ser Ser
 265 270 275
 5 gtg tac tgaggcccg cagctctggc cacagggacc tctgcagtgc cccctaagtg 1046
 Val Tyr
 acccgacac ttccgagggg gccatcacgc cctgtgtata taacgtttcc ggtattactc 1106
 tgctacacgt agccttttta cttttggggg tttgtttttg ttctgaactt tcctgttacc 1166
 ttttcagggc tgacgtcaca tgtaggtggc gtgtatgagt ggagacgggc ctgggtcttg 1226
 10 gggactggag ggcaggggtc cttctgccct ggggtcccag ggtgctctgc ctgctcagcc 1286
 aggcctctcc tgggagccac tcgcccagag actcagcttg gccaaacttg ggggctgtgt 1346
 ccaccagcc cgcccgctct gtgggctgca cagctcacct tgttccctcc tgccccggtt 1406
 cgagagccga gtctgtgggc actctctgcc ttcatgcacc tgtcctttct aacacgtcgc 1466
 cttcaactgt aatcacaaca tcctgactcc gtcatttaat aaagaaggaa catcaggcat 1526
 15 gcaaaaaaaaa aaaaaaaaaa 1545

 <210> 156
 <211> 1068
 <212> DNA
 20 <213> Homo sapiens

 <220>
 <221> CDS
 <222> 11..529
 25
 <400> 156
 gaagcacgga atg tgt ctc ctg ctg ggg gcc acg ggc gtc ggg aag acg 49
 Met Cys Leu Leu Leu Gly Ala Thr Gly Val Gly Lys Thr
 1 5 10
 30 ctg ctg gtg aaa cgg ctg cag gag gtg agc tcc cgg gat ggg aaa ggc 97
 Leu Leu Val Lys Arg Leu Gln Glu Val Ser Ser Arg Asp Gly Lys Gly
 15 20 25
 gac ctg ggg gag ccg ccc ccg aca cgg ccc acg gtg ggc acc aat ctt 145
 Asp Leu Gly Glu Pro Pro Pro Thr Arg Pro Thr Val Gly Thr Asn Leu
 35 30 35 40 45
 act gac atc gtg gca cag aga aag atc acc atc cgg gag ctt ggg ggg 193
 Thr Asp Ile Val Ala Gln Arg Lys Ile Thr Ile Arg Glu Leu Gly Gly
 50 55 60
 tgc atg ggc ccc atc tgg tcc agt tac tat gga aac tgc cgt tct ctc 241
 Cys Met Gly Pro Ile Trp Ser Ser Tyr Tyr Gly Asn Cys Arg Ser Leu
 65 70 75
 ctg ttt gtg atg gac gcc tct gac ccc acc cag ctc tct gca tcc tgt 289
 Leu Phe Val Met Asp Ala Ser Asp Pro Thr Gln Leu Ser Ala Ser Cys
 80 85 90
 45 gtg cag ctc tta ggt ctc ctt tct gca gaa caa ctt gca gaa gca tgc 337
 Val Gln Leu Leu Gly Leu Leu Ser Ala Glu Gln Leu Ala Glu Ala Ser
 95 100 105
 gtg ctg ata ctc ttc aat aaa atc gac cta ccc tgt tac atg tcc acg 385
 Val Leu Ile Leu Phe Asn Lys Ile Asp Leu Pro Cys Tyr Met Ser Thr
 110 115 120 125
 50 gag gag atg aag tca tta atc agg ctt cca gac atc att gct tgt gcc 433
 Glu Glu Met Lys Ser Leu Ile Arg Leu Pro Asp Ile Ile Ala Cys Ala
 130 135 140
 aag cag aac atc acc acg gca gaa atc agc gcc cgt gaa ggc act ggc 481
 Lys Gln Asn Ile Thr Thr Ala Glu Ile Ser Ala Arg Glu Gly Thr Gly
 145 150 155
 tta gca ggg gtg ctg gcc tgg ctc cag gcc acc cac aga gcc aac gat 529
 Leu Ala Gly Val Leu Ala Trp Leu Gln Ala Thr His Arg Ala Asn Asp
 160 165 170
 60 tgactgcacg gcagagggcg agctggcctg agctggggag aggtggcaga gggcagtatg 589
 gctttgctgc caatagtttc ttctcacagg ggcagaataa cccaaagtaa ccctacatga 649
 tggggctctg tgctgggatg caatgatgtg taaactgagg catgtggaga tggaagtga 709
 catctggcct ctgaaaaaag tgtccccagg ggctaggcat ggtggctcac acctgtaatc 769

ccagcacttt gagaggccga ggcgggtgta tcacctgagg tcgggagttc gagactagcc 829
 tgaccaacat ggagaaaccc tgtctctact aaaaatacaa aattagctgg gtgtgctgg 889
 gcatgcctgt aatctcagct acttgggagg ctgagacagg agaatccctt gaacctggga 949
 ggtggagggt gcagtgaagc gagatcatgc cattgcactg cacctgggca acaagagtga 1009
 5 aactccgtct taaaaaatat aagaaataaa aaaataaaaa cctaaaaaaa aaaaaaaaaa 1068

<210> 157

<211> 1097

<212> DNA

10 <213> Homo sapiens

<220>

<221> CDS

<222> 135..749

15

<400> 157

aacgaaacgg taaccagccc tgggaagccc gcaagaggcc tcagcgggtgg ccgtccgagc 60
 gccgagagggt gaggggtgccc ccgcctcacc tgcagagggg ccgttccggg ctcgaacccg 120
 gcaccttccg gaaa atg gcg gct gcc agg ccc agc ctg ggc mga gtc ctc 170
 20 Met Ala Ala Ala Arg Pro Ser Leu Gly Arg Val Leu
 1 5 10
 cca gga tcc tct gtc ctg ttc ctg tgt gac atg cag gag aag ttc cgc 218
 Pro Gly Ser Ser Val Leu Phe Leu Cys Asp Met Gln Glu Lys Phe Arg
 15 20 25
 25 cac aac atc gcc tac ttc cca cag atc gtc tca gtg gct gcc cgc atg 266
 His Asn Ile Ala Tyr Phe Pro Gln Ile Val Ser Val Ala Ala Arg Met
 30 35 40
 ctc aag gtg gcc cgg ctg ctt gag gtg cca gtc atg ctg acg gag cag 314
 Leu Lys Val Ala Arg Leu Leu Glu Val Pro Val Met Leu Thr Glu Gln
 30 45 50 55 60
 tac cca caa ggc ctg ggc ccc acg gtg ccc gag ctg ggg act gag ggc 362
 Tyr Pro Gln Gly Leu Gly Pro Thr Val Pro Glu Leu Gly Thr Glu Gly
 65 70 75
 ctt cgg ccg ctg gcc aag acc tgc ttc agc atg gtg cct gcc ctg cag 410
 35 Leu Arg Pro Leu Ala Lys Thr Cys Phe Ser Met Val Pro Ala Leu Gln
 80 85 90
 cag gag ctg gac agt cgg ccc cag ctg cgc tct gtg ctg ctc tgt ggc 458
 Gln Glu Leu Asp Ser Arg Pro Gln Leu Arg Ser Val Leu Cys Gly
 95 100 105
 40 att gag gca cag gcc tgc atc ttg aac acg acc ctg gac ctc cta gac 506
 Ile Glu Ala Gln Ala Cys Ile Leu Asn Thr Thr Leu Asp Leu Leu Asp
 110 115 120
 cgg ggg ctg cag gtc cat gtg gtg gtg gac gcc tgc tcc tca cgc agc 554
 Arg Gly Leu Gln Val His Val Val Val Asp Ala Cys Ser Ser Arg Ser
 45 125 130 135 140
 cag gtg gac cgt ctg gtg gct ctg gcc cgc atg aga cag agt ggt gcc 602
 Gln Val Asp Arg Leu Val Ala Leu Ala Arg Met Arg Gln Ser Gly Ala
 145 150 155
 ttc ctc tcc acc agc gaa ggg ctc att ctg cag ctt gtg ggc gat gcc 650
 50 Phe Leu Ser Thr Ser Glu Gly Leu Ile Leu Gln Leu Val Gly Asp Ala
 160 165 170
 gtc cac ccc cag ttc aag gag atc cag aaa ctc atc aag gag ccc gcc 698
 Val His Pro Gln Phe Lys Glu Ile Gln Lys Leu Ile Lys Glu Pro Ala
 175 180 185
 55 cca gac agc gga ctg ctg ggc ctc ttc caa ggc cag aac tcc ctc ctc 746
 Pro Asp Ser Gly Leu Leu Gly Leu Phe Gln Gly Gln Asn Ser Leu Leu
 190 195 200
 cac tgaactccaa cctgccttg agggaagacc accctcctgt caccgggacc 799
 His
 60 205
 tcagtggaag cccgttcccc ccatccctgg atcccaagag tgggtgcgatc caccaggagt 859
 gccgccccct tgtggggggg ggcagggtgc tgccttccca ttggacagct gctccggaa 919
 atgcaaatga gactcctgga aactgggtgg gaattggctg agccaagatg gaggcggggc 979

tcggccccgg gccacttcac ggggcgggaa ggggagggga agaagagtct cagactgtgg 1039
gacacggact cgcagaataa acatatatgt ggctgtggac caaaaaaaaa aaaaaaaa 1097

<210> 158
5 <211> 894
<212> DNA
<213> Homo sapiens

<220>
10 <221> CDS
<222> 98..637

<400> 158
ctttggggcc gaagtgggcg tgcggctcgc gctgttcgcg gccttcctgg tgacggagct 60
15 gctccccccg ttccagagac tcatccagcc ggaggag atg tgg ctc tac cgg aac 115
Met Trp Leu Tyr Arg Asn
1 5

ccc tac gtg gag gcg gag tat ttc ccc acc aag ccg atg ttt gtt att 163
Pro Tyr Val Glu Ala Glu Tyr Phe Pro Thr Lys Pro Met Phe Val Ile
20 10 15 20

gca ttt ctc tct cca ctg tct ctg atc ttc ctg gcc aaa ttt ctc aag 211
Ala Phe Leu Ser Pro Leu Ser Leu Ile Phe Leu Ala Lys Phe Leu Lys
25 30 35

aag gca gac aca aga gac agc aga caa gcc tgc ctg gct gcc agc ctt 259
25 Lys Ala Asp Thr Arg Asp Ser Arg Gln Ala Cys Leu Ala Ala Ser Leu
40 45 50

gcc ctg gct ctg aat ggc gtc ttt acc aac aca ata aaa ctg atc gta 307
Ala Leu Ala Leu Asn Gly Val Phe Thr Asn Thr Ile Lys Leu Ile Val
55 60 65 70

ggg agg cca cgc cca gat ttc ttc tac cgc tgc ttc cct gat ggg cta 355
30 Gly Arg Pro Arg Pro Asp Phe Phe Tyr Arg Cys Phe Pro Asp Gly Leu
75 80 85

gcc cat tct gac ttg atg tgt aca ggg gat aag gac gtg gtg aat gag 403
35 Ala His Ser Asp Leu Met Cys Thr Gly Asp Lys Asp Val Val Asn Glu
90 95 100

ggc cga aag agc ttc ccc agt gga cat tct tcc ttt gca ttt gct ggt 451
Gly Arg Lys Ser Phe Pro Ser Gly His Ser Ser Phe Ala Phe Ala Gly
105 110 115

ctg gcc ttt gcg tcc ttc tac ctg gca ggg aag tta cac tgc ttc aca 499
40 Leu Ala Phe Ala Ser Phe Tyr Leu Ala Gly Lys Leu His Cys Phe Thr
120 125 130

cca caa ggc cgt ggg aaa tct tgg agg ttc tgt gcc ttt ctg tca cct 547
Pro Gln Gly Arg Gly Lys Ser Trp Arg Phe Cys Ala Phe Leu Ser Pro
135 140 145 150

cta ctt ttt gca gct gtg att gca ctg tcc cgc aca tgt gac tac aag 595
45 Leu Leu Phe Ala Val Ile Ala Leu Ser Arg Thr Cys Asp Tyr Lys
155 160 165

cat cac tgg caa gat ctg ctc aaa tgc acc aac act gcc aag 637
His His Trp Gln Asp Leu Leu Lys Cys Thr Asn Thr Ala Lys
50 170 175 180

tgactaaggt agaaaagaaa aatgacaggt atcgatcatct gaaggacaga tgaatctttt 697
tctgcccctt cttcacaatg gaatataagg aacaattatg ggatgtcatc agaatggatg 757
ccataggacc tacagctccc tttctcttta ttgtgattat acttttaaata tgacattgtc 817
ttttatgtgt atgttcctat attttcaatg tatctttttc cttcagtaaa cctgatattc 877
55 aaaaaaaaa aaaaaaa 894

<210> 159
<211> 703
<212> DNA
60 <213> Homo sapiens

<220>
<221> CDS

<222> 221..670

<400> 159

5 aaggaagcgc cgcccccttcc tacggctacg ggaaggatcg tccagtggct gaggctggac 60
 agagcacaga caaaagataa aaagcaagat ttgagagagg ttcttggatc aactgggtca 120
 atctgcttag ttctacaaag tggagtttct gggcatcatt cttcatttct gtacacaaag 180
 tgctgtgaag ctcaagaaga aatagctctg cacaggaacg atg tgc act gcc cta 235
 Met Cys Thr Ala Leu
 1 5
 10 ctg ctt ctt tat cta aga tgg tgt ttc aac tta aaa ctt gtg aat gtg 283
 Leu Leu Leu Tyr Leu Arg Trp Cys Phe Asn Leu Lys Leu Val Asn Val
 10 15 20
 aaa tat gag cca aaa gac tct ctc ggc cct gaa atg acc ttt gta gca 331
 Lys Tyr Glu Pro Lys Asp Ser Leu Gly Pro Glu Met Thr Phe Val Ala
 25 30 35
 15 gat gct gcc aga ggc ccc ctg tta tcc tcc ctg gac tct cca gct aac 379
 Asp Ala Ala Arg Gly Pro Leu Leu Ser Ser Leu Asp Ser Pro Ala Asn
 40 45 50
 20 ctg atg tca act gcc agt gtg tgc atc tcc tta cct gag ggc tgt tct 427
 Leu Met Ser Thr Ala Ser Val Cys Ile Ser Leu Pro Glu Gly Cys Ser
 55 60 65
 ggt ggc agg agt cct tgc tac tca cag aaa tgg cca cca gaa gtg cca 475
 Gly Gly Arg Ser Pro Cys Tyr Ser Gln Lys Trp Pro Pro Glu Val Pro
 70 75 80 85
 25 gaa aaa tta acc tcc ctt ggc cag cag tcc tca acc agc tcc ctc act 523
 Glu Lys Leu Thr Ser Leu Gly Gln Gln Ser Ser Thr Ser Ser Leu Thr
 90 95 100
 gac act gat gtg cag gtg tct cct atg ctg gtt gct gga gtc aac cac 571
 Asp Thr Asp Val Gln Val Ser Pro Met Leu Val Ala Gly Val Asn His
 105 110 115
 30 agc agc agc ctt ctt gac aac ata ccc ttc act ggc tgc ctt cct ttc 619
 Ser Ser Ser Leu Leu Asp Asn Ile Pro Phe Thr Gly Cys Leu Pro Phe
 120 125 130
 cat ctc tct tct tca ctc ccc tac cta tgt ctc cta ggc tct ccc ttc 667
 His Leu Ser Ser Ser Leu Pro Tyr Leu Cys Leu Leu Gly Ser Pro Phe
 135 140 145
 aaa taaacagctt gcacttgaaa aaaaaaaaaa aaa 703
 Lys
 150

40

<210> 160

<211> 849

<212> DNA

<213> Homo sapiens

45

<220>

<221> CDS

<222> 165..674

50 <400> 160

50 aaaactgagg cctgggagca ggaacctgta ggcagcgctt gagggtagcg ggatagcagc 60
 tgcaagcgcg cgtgggagggc gggggctctg ggcggaacaa aaatcacagg atgtcagagg 120
 atgtttcccg ggaagaactg ggataaagga aggggtcccg cacc atg gag gac ccg 176
 Met Glu Asp Pro
 1
 55 aac cct gaa gag aac atg aag cag cag gat tca ccc aag gag aga agt 224
 Asn Pro Glu Glu Asn Met Lys Gln Gln Asp Ser Pro Lys Glu Arg Ser
 5 10 15 20
 ccc cag agc cca gga ggc aac atc tgc cac ctg ggg gcc ccg aag tgc 272
 60 Pro Gln Ser Pro Gly Gly Asn Ile Cys His Leu Gly Ala Pro Lys Cys
 25 30 35
 acc cgc tgc ctc atc acc ttc gca gat tcc aag ttc cag gag cgt cac 320
 Thr Arg Cys Leu Ile Thr Phe Ala Asp Ser Lys Phe Gln Glu Arg His

			40				45				50								
	atg	aag	cgg	gag	cac	cca	gcg	gac	ttc	gtg	gcc	cag	aag	ctg	cag	ggg			368
	Met	Lys	Arg	Glu	His	Pro	Ala	Asp	Phe	Val	Ala	Gln	Lys	Leu	Gln	Gly			
			55					60					65						
5	gtc	ctc	ttc	atc	tgc	ttc	acc	tgc	gcc	cgc	tcc	ttc	ccc	tcc	tcc	aaa			416
	Val	Leu	Phe	Ile	Cys	Phe	Thr	Cys	Ala	Arg	Ser	Phe	Pro	Ser	Ser	Lys			
			70					75					80						
	gcc	cta	atc	acc	cac	cag	cgc	agc	cac	ggt	cca	gcc	gcc	aag	ccc	acc			464
	Ala	Leu	Ile	Thr	His	Gln	Arg	Ser	His	Gly	Pro	Ala	Ala	Lys	Pro	Thr			
10	85					90					95					100			
	ctg	ccg	gtt	gca	acc	act	act	gcc	cag	ccc	acc	ttc	cct	tgt	cct	gac			512
	Leu	Pro	Val	Ala	Thr	Thr	Thr	Ala	Gln	Pro	Thr	Phe	Pro	Cys	Pro	Asp			
					105					110					115				
	tgt	ggc	aag	acc	ttt	ggg	cag	gct	gtt	tct	ctg	agg	cgg	cac	cgc	cag			560
15	Cys	Gly	Lys	Thr	Phe	Gly	Gln	Ala	Val	Ser	Leu	Arg	Arg	His	Arg	Gln			
				120				125						130					
	atg	cat	gag	gtc	cgt	gcc	cct	cct	ggc	acc	ttc	gcc	tgc	aca	gag	tgc			608
	Met	His	Glu	Val	Arg	Ala	Pro	Pro	Gly	Thr	Phe	Ala	Cys	Thr	Glu	Cys			
				135				140					145						
20	ggc	cag	gac	ttt	gct	cag	gaa	gca	ggg	ctg	cat	caa	cac	tac	att	cgg			656
	Gly	Gln	Asp	Phe	Ala	Gln	Glu	Ala	Gly	Leu	His	Gln	His	Tyr	Ile	Arg			
		150				155				160									
	cat	gcc	cgg	ggg	gag	ctc	tgagtgcagc	ttaagcctct	ccacggtgac										704
	His	Ala	Arg	Gly	Glu	Leu													
25	165					170													
	gggtggtctct	gtggctggta	ggactcaccc	atgatatggg	gtgcaggaac	tctggggggcc													764
	ctgaaggatt	tgcttcctc	ccctgggaag	gcagagggct	cttaataaag	aggaccacga													824
	agattctcaa	aaaaaaaaaa	aaaaa																849
30	<210>	161																	
	<211>	846																	
	<212>	DNA																	
	<213>	Homo sapiens																	
35	<220>																		
	<221>	CDS																	
	<222>	165..671																	
	<400>	161																	
40	aaaactgagg	cctgggagca	ggaacctgta	ggcagcgctt	gagggtagcg	ggatagcagc													60
	tgcaagcgcg	cgtgggaggc	gggggctctg	ggcggaaaca	aaatcacagg	atgtcagagg													120
	atgtttcccg	ggaagaactg	ggataaagga	aggggtcccg	cacc	atg	gag	gac	ccg										176
					Met	Glu	Asp	Pro											
					1														
45	aac	cct	gaa	gag	aac	atg	aag	cag	cag	gat	tca	ccc	aag	gag	aga	agt			224
	Asn	Pro	Glu	Glu	Asn	Met	Lys	Gln	Gln	Asp	Ser	Pro	Lys	Glu	Arg	Ser			
	5					10					15				20				
	ccc	cag	ccc	agg	agg	caa	cat	ctg	cca	cct	ggg	ggc	ccc	gaa	gtg	cac			272
	Pro	Gln	Pro	Arg	Arg	Gln	His	Leu	Pro	Pro	Gly	Gly	Pro	Glu	Val	His			
50					25					30					35				
	ccg	ctg	cct	cat	cac	ctt	cgc	aga	ttc	caa	gtt	cca	gga	gcg	tca	cat			320
	Pro	Leu	Pro	His	His	Leu	Arg	Arg	Phe	Gln	Val	Pro	Gly	Ala	Ser	His			
					40				45					50					
	gaa	gcg	gga	gca	ccc	agc	gga	ctt	cgt	ggc	cca	gaa	gct	gca	ggg	ggt			368
55	Glu	Ala	Gly	Ala	Pro	Ser	Gly	Leu	Arg	Gly	Pro	Glu	Ala	Ala	Gly	Gly			
		55				60				65									
	cct	ctt	cat	ctg	ctt	cac	ctg	cgc	ccg	ctc	ctt	ccc	ctc	ctc	caa	agc			416
	Pro	Leu	His	Leu	Leu	His	Leu	Arg	Pro	Leu	Leu	Pro	Leu	Leu	Gln	Ser			
		70				75				80									
60	cct	aat	cac	cca	cca	gcg	cag	cac	ggt	cca	gcc	gcc	aag	ccc	acc	ctg			464
	Pro	Asn	His	Pro	Pro	Ala	Gln	His	Gly	Pro	Ala	Ala	Lys	Pro	Thr	Leu			
	85					90				95					100				
	ccg	gtt	gca	acc	act	act	gcc	cag	ccc	acc	ttc	cct	tgt	cct	gac	tgt			512

	Pro	Val	Ala	Thr	Thr	Thr	Ala	Gln	Pro	Thr	Phe	Pro	Cys	Pro	Asp	Cys	
					105					110					115		
	ggc	aag	acc	ttt	ggg	cag	gct	gtt	tct	ctg	agg	cgg	cac	cgc	cag	atg	560
5	Gly	Lys	Thr	Phe	Gly	Gln	Ala	Val	Ser	Leu	Arg	Arg	His	Arg	Gln	Met	
				120					125					130			
	cat	gag	gtc	cgt	gcc	cct	cct	ggc	acc	ttc	gcc	tgc	aca	gag	tgc	ggg	608
	His	Glu	Val	Arg	Ala	Pro	Pro	Gly	Thr	Phe	Ala	Cys	Thr	Glu	Cys	Gly	
				135				140					145				
	cag	gac	ttt	gct	cag	gaa	gca	ggg	ctg	cat	caa	cac	tac	att	cgg	cat	656
10	Gln	Asp	Phe	Ala	Gln	Glu	Ala	Gly	Leu	His	Gln	His	Tyr	Ile	Arg	His	
		150					155					160					
	gcc	cgg	ggg	gag	ctc	tgagtgcagc	ttaagcctct	ccacgggtgac	gggtggctct								711
	Ala	Arg	Gly	Glu	Leu												
				165													
15	gtggctggtg	ggactcaccc	atgatatggg	gtgcaggaac	tctggggggcc	ctgaaggatt											771
	tgcttccttc	ccctgggaag	gcagagggct	cttaataaag	aggacccaga	agattctcaa											831
	aaaaaaaaaa	aaaaaa															846
	<210> 162																
20	<211> 1176																
	<212> DNA																
	<213> Homo sapiens																
	<220>																
25	<221> CDS																
	<222> 28..1128																
	<400> 162																
30	ctttcctgcc	tctgattccg	ggctgtc	atg	gcg	acc	ccc	aac	aat	ctg	acc	ccc					54
				Met	Ala	Thr	Pro	Asn	Asn	Leu	Thr	Pro					
				1				5									
	acc	aac	tgc	agc	tgg	tgg	ccc	atc	tcc	gcg	ctg	gag	agc	gat	gcg	gcc	102
	Thr	Asn	Cys	Ser	Trp	Trp	Pro	Ile	Ser	Ala	Leu	Glu	Ser	Asp	Ala	Ala	
	10				15					20					25		
35	aag	cca	gcg	gag	gcc	ccc	gac	gct	ccc	gag	gcg	gcc	agc	ccc	gcc	cat	150
	Lys	Pro	Ala	Glu	Ala	Pro	Asp	Ala	Pro	Glu	Ala	Ala	Ser	Pro	Ala	His	
				30					35					40			
	tgg	ccc	agg	gag	agc	ctg	gtt	ctg	tac	cac	tgg	acc	cag	tcc	ttc	agc	198
	Trp	Pro	Arg	Glu	Ser	Leu	Val	Leu	Tyr	His	Trp	Thr	Gln	Ser	Phe	Ser	
40				45				50					55				
	tcg	cag	aag	gtg	cgg	ctg	gtg	atc	gcc	gag	aag	ggc	ctg	gtg	tgc	gag	246
	Ser	Gln	Lys	Val	Arg	Leu	Val	Ile	Ala	Glu	Lys	Gly	Leu	Val	Cys	Glu	
			60				65					70					
	gag	cgg	gac	gtg	agc	ctg	cca	cag	agc	gag	cac	aag	gag	ccc	tgg	ttc	294
45	Glu	Arg	Asp	Val	Ser	Leu	Pro	Gln	Ser	Glu	His	Lys	Glu	Pro	Trp	Phe	
		75				80					85						
	atg	cgg	ctc	aac	ctg	ggc	gag	gag	gtg	ccc	gtc	atc	atc	cac	cgc	gac	342
	Met	Arg	Leu	Asn	Leu	Gly	Glu	Glu	Val	Pro	Val	Ile	Ile	His	Arg	Asp	
	90				95					100						105	
50	aac	atc	atc	agt	gac	tat	gac	cag	atc	att	gac	tat	gtg	gag	cgc	acc	390
	Asn	Ile	Ile	Ser	Asp	Tyr	Asp	Gln	Ile	Ile	Asp	Tyr	Val	Glu	Arg	Thr	
				110						115					120		
	ttc	aca	gga	gag	cac	gtg	gtg	gcc	ctg	atg	ccc	gag	gtg	ggc	agc	ctg	438
	Phe	Thr	Gly	Glu	His	Val	Val	Ala	Leu	Met	Pro	Glu	Val	Gly	Ser	Leu	
55				125					130					135			
	cag	cac	gca	cgg	gtg	ctg	cag	tac	cgg	gag	ctg	ctg	gac	gca	ctg	ccc	486
	Gln	His	Ala	Arg	Val	Leu	Gln	Tyr	Arg	Glu	Leu	Leu	Asp	Ala	Leu	Pro	
			140					145					150				
	atg	gat	gcc	tac	acg	cat	ggc	tgc	atc	ctg	cat	ctc	gag	ctc	acc	acc	534
60	Met	Asp	Ala	Tyr	Thr	His	Gly	Cys	Ile	Leu	His	Leu	Glu	Leu	Thr	Thr	
		155					160					165					
	gac	tcc	atg	atc	ccc	aag	tac	gcc	acg	gcc	gag	atc	cgc	aga	cat	tta	582
	Asp	Ser	Met	Ile	Pro	Lys	Tyr	Ala	Thr	Ala	Glu	Ile	Arg	Arg	His	Leu	

	170					175					180						185	
	gcc aat gcc acc acg gac ctc atg aaa ctg gac cat gaa gag gag ccc																	630
	Ala Asn Ala Thr Thr Asp Leu Met Lys Leu Asp His Glu Glu Glu Pro																	
					190						195						200	
5	cag ctc tcc gag ccc tac ctt tct aaa caa aag aag ctc atg gcc aag																	678
	Gln Leu Ser Glu Pro Tyr Leu Ser Lys Gln Lys Lys Leu Met Ala Lys																	
				205							210						215	
	atc ttg gag cat gat gat gtg agc tac ctg aag aag atc ctc ggg gaa																	726
	Ile Leu Glu His Asp Asp Val Ser Tyr Leu Lys Lys Ile Leu Gly Glu																	
10			220															
	ctg gcc atg gtg ctg gac cag att gag gcg gag ctg gag aag agg aag																	774
	Leu Ala Met Val Leu Asp Gln Ile Glu Ala Glu Leu Glu Lys Arg Lys																	
			235															
	ctg gag aac gag ggg cag aaa tgc gag ctg tgg ctc tgt ggc tgt gcc																	822
15	Leu Glu Asn Glu Gly Gln Lys Cys Glu Leu Trp Leu Cys Gly Cys Ala																	
	250					255												265
	ttc acc ctc gct gat gtc ctc ctg gga gcc acc ctg cac cgc ctc aag																	870
	Phe Thr Leu Ala Asp Val Leu Leu Gly Ala Thr Leu His Arg Leu Lys																	
				270														
20	ttc ctg gga ctg tcc aag aaa tac tgg gaa gat ggc agc cgg ccc aac																	918
	Phe Leu Gly Leu Ser Lys Lys Tyr Trp Glu Asp Gly Ser Arg Pro Asn																	
			285															
	ctg cag tcc ttc ttt gag agg gtc cag aga cgc ttt gcc ttc cgg aaa																	966
	Leu Gln Ser Phe Phe Glu Arg Val Gln Arg Arg Phe Ala Phe Arg Lys																	
25			300															
	gtc ctg ggt gac atc cac acc acc ctg ctg tcg gcc gtc atc ccc aat																	1014
	Val Leu Gly Asp Ile His Thr Thr Leu Leu Ser Ala Val Ile Pro Asn																	
			315															
	gct ttc cgg ctg gtc aag agg aaa ccc cca tcc ttc ttc ggg gcg tcc																	1062
30	Ala Phe Arg Leu Val Lys Arg Lys Pro Pro Ser Phe Phe Gly Ala Ser																	
	330					335												345
	ttc ctc atg ggc tcc ctg ggt ggg atg ggc tac ttt gcc tac tgg tac																	1110
	Phe Leu Met Gly Ser Leu Gly Gly Met Gly Tyr Phe Ala Tyr Trp Tyr																	
				350														360
35	ctc aag aaa aaa tac atc tagggccagg cctggggcctt ggtgtctgac																	1158
	Leu Lys Lys Lys Tyr Ile																	
				365														
	aaaaaamaaa aaaaaaaa																	1176
40	<210> 163																	
	<211> 1084																	
	<212> DNA																	
	<213> Homo sapiens																	
45	<220>																	
	<221> CDS																	
	<222> 135..194																	
	<400> 163																	
50	aacgaaacgg taaccagccc tgggaagccc gcaagaggcc tcagcgggtgg ccgtccgagc																	60
	gccgagaggt gaggggtgcc ccgcctcacc tgcagagggg ccgttcgggg ctcgaacccg																	120
	gcaccttcgg gaaa atg gcg gct gcc agg ccc agc ctg ggc cga gtc ctc																	170
				Met	Ala	Ala	Ala	Arg	Pro	Ser	Leu	Gly	Arg	Val	Leu			
				1				5						10				
55	cca gga tcc tct cct gtt cct gtg tgacatgcag gagaagttcc gccacaacat																	224
	Pro Gly Ser Ser Pro Val Pro Val																	
				15				20										
	cgctacttc ccacagatcg tctcagtggc tgcccgcgatg ctcaaggtgg cccggctgct																	284
	tgaggtgcc gtcagtctga cggagcagta cccacaaggc ctggggcccca cggtgcccg																	344
60	gctggggact gagggccttc ggccgctggc caagacctgc ttcagcatgg tgccctgcct																	404
	gcagcaggag ctggacagtc ggccccagct gcgctctgtg ctgctctgtg gcattgaggc																	464
	acaggcctgc atcttgaaca cgaccttgga cctcctagac cgggggctgc aggtccatgt																	524
	qqtqqtqgac gcctgtcct cacgcagcca ggtggaccgg ctggtggctc tggcccgc																	584

	gagacagagt	ggtgccttcc	tctccaccag	cgaagggctc	attctgcagc	ttgtgggcca	644
	tgccgtccac	ccccagttca	aggagatcca	gaaactcatc	aaggagcccg	ccccagacag	704
	cggactgctg	ggcctcttcc	aaggccagaa	ctccctcctc	cactgaactc	caaccctgcc	764
	ttgaggggaag	accaccctcc	tgtcaccg	acctcagtgg	aagcccgttc	cccccatccc	824
5	tggatcccaa	gagtgggtgcg	atccaccagg	agtgccgccc	ccttgggggg	ggcaggggtgc	884
	tgccttccca	ttggacagct	gctcccggaa	atgcaaatga	gactcctgga	aactgggtgg	944
	gaattggctg	agccaagatg	gaggcggggc	tcggcccccg	gccacttcac	ggggcgggaa	1004
	ggggagggga	agaagagtct	cagactgtgg	gacacggact	cgcagaataa	acatatatgt	1064
	ggcaaaaaaa	aaaaaaaaaa					1084
10							
	<210> 164						
	<211> 1793						
	<212> DNA						
	<213> Homo sapiens						
15							
	<220>						
	<221> CDS						
	<222> 173..847						
20							
	<400> 164						
	gsmggrggcc	attacctaga	acatcsta	cgaarratta	tttgaaaaac	cactgggttc	60
	cgagttcatt	actacaggaa	aaactgttct	cttctgtggc	acagagaacc	ctgcttcaaa	120
	gcagaagtag	cagttccgga	gtccagctgg	ctaaaactca	tcccagagga	ta atg gca	178
						Met Ala	
25						1	
	acc cat gcc tta gaa atc gct ggg ctg ttt ctt ggt ggt gtt gga atg	226					
	Thr His Ala Leu Glu Ile Ala Gly Leu Phe Leu Gly Gly Val Gly Met						
	5 10 15						
	gtg ggc aca gtg gct gtc act gtc atg cct cag tgg aga gtg tcg gcc	274					
30	Val Gly Thr Val Ala Val Thr Val Met Pro Gln Trp Arg Val Ser Ala						
	20 25 30						
	ttc att gaa aac aac atc gtg gtt ttt gaa aac ttc tgg gaa gga ctg	322					
	Phe Ile Glu Asn Asn Ile Val Val Phe Glu Asn Phe Trp Glu Gly Leu						
	35 40 45 50						
35	tgg atg aat tgc gtg agg cag gct aac atc agg atg cag tgc aaa atc	370					
	Trp Met Asn Cys Val Arg Gln Ala Asn Ile Arg Met Gln Cys Lys Ile						
	55 60 65						
	tat gat tcc ctg ctg gct ctt tct ccg gac cta cag gca gcc aga gga	418					
	Tyr Asp Ser Leu Leu Ala Leu Ser Pro Asp Leu Gln Ala Ala Arg Gly						
40	70 75 80						
	ctg atg tgt gct gct tcc gtg atg tcc ttc ttg gct ttc atg atg gcc	466					
	Leu Met Cys Ala Ala Ser Val Met Ser Phe Leu Ala Phe Met Met Ala						
	85 90 95						
	atc ctt ggc atg aaa tgc acc agg tgc acg ggg gac aat gag aag gtg	514					
45	Ile Leu Gly Met Lys Cys Thr Arg Cys Thr Gly Asp Asn Glu Lys Val						
	100 105 110						
	aag gct cac att ctg ctg acg gct gga atc atc ttc atc atc acg ggc	562					
	Lys Ala His Ile Leu Leu Thr Ala Gly Ile Ile Phe Ile Ile Thr Gly						
	115 120 125 130						
50	atg gtg gtg ctc atc cct gtg agc tgg gtt gcc aat gcc atc atc aga	610					
	Met Val Val Leu Ile Pro Val Ser Trp Val Ala Asn Ala Ile Ile Arg						
	135 140 145						
	gat ttc tat aac tca ata gtg aat gtt gcc caa aaa cgt gag ctt gga	658					
	Asp Phe Tyr Asn Ser Ile Val Asn Val Ala Gln Lys Arg Glu Leu Gly						
55	150 155 160						
	gaa gct ctc tac tta gga tgg acc acg gca ctg gtg ctg att gtt gga	706					
	Glu Ala Leu Tyr Leu Gly Trp Thr Thr Ala Leu Val Leu Ile Val Gly						
	165 170 175						
	gga gct ctg ttc tgc tgc gtt ttt tgt tgc aac gaa aag agc agt agc	754					
60	Gly Ala Leu Phe Cys Cys Val Phe Cys Cys Asn Glu Lys Ser Ser Ser						
	180 185 190						
	tac aga tac tcg ata cct tcc cat cgc aca acc caa aaa agt tat cac	802					
	Tyr Arg Tyr Ser Ile Pro Ser His Arg Thr Thr Gln Lys Ser Tyr His						

	195	200	205	210	
	acc gga aag aag tca ccg agc gtc tac tcc aga agt cag tat gtg				847
	Thr Gly Lys Lys Ser Pro Ser Val Tyr Ser Arg Ser Gln Tyr Val				
		215	220	225	
5	tagttgtgta tgttttttta acttttactat aaagccatgc aaatgacaaa aatctatatatt				907
	acttttctcaa aatggacccc aaagaaactt tgattttactg ttcttaactg cctaactctta				967
	attacaggaa ctgtgcatca gctattttatg attctataag ctatttcagc agaatgagat				1027
	attaaaccca atgctttgat tggtctagaa agtattgtaa tttgttttct aaggtggttc				1087
	aagcatctac tcttttttattc atttacttca aaatgacatt gctaaagact gcattattct				1147
10	actactgtaa tttctccacg acatagcatt atgtacatag atgagtgtaa catttatatc				1207
	tcacatagag acatgcttat atgggttttat ttaaaatgaa atgccagtcc attacactga				1267
	ataaatagaa ctcaactatt gcttttcagg gaaatcatgg ataggggtga agaaggttac				1327
	tattaattgt ttaaaaacag cttatggatt aatgtcctcc atttataatg aagattaaaa				1387
	tgaaggcttt aatcagcatt gtaaaggaaa ttgaatggct ttctgatatg ctgtttttta				1447
15	gcctaggagt tagaaatcct aacttcttta tctcttctc ccagaggctt ttttttctt				1507
	gtgtattaaa ttaacatttt taaaaagcag atattttgtc aaggggcttt gcattcaaac				1567
	tgcttttcca gggctatact cagaagaaa ataaaaagtg gatctaagaa aaagtgatgg				1627
	ttttaggaaa gtgaaaatat ttttgtttt gtatttgaag aagaatgatg cattttgaca				1687
	agaaatcata tatgtatgta tatattttta taagtatttg agtacagact ttgaggtttc				1747
20	atcaatataa ataaaagagc agaaaagtaa aaaaaaaaa aaaaaa				1793
	<210> 165				
	<211> 1849				
	<212> DNA				
25	<213> Homo sapiens				
	<220>				
	<221> CDS				
	<222> 8..1141				
30	<220>				
	<221> misc_feature				
	<222> 1707				
	<223> n=a, g, c or t				
35	<400> 165				
	cgttgcc atg gat cct ggg gac gac tgg ctg gtg gaa tcc ttg cgc ttg				49
	Met Asp Pro Gly Asp Asp Trp Leu Val Glu Ser Leu Arg Leu				
	1 5 10				
40	tac cag gat ttc tat gca ttc gac ctg tca gga gcc act cga gtc ctt				97
	Tyr Gln Asp Phe Tyr Ala Phe Asp Leu Ser Gly Ala Thr Arg Val Leu				
	15 20 25 30				
	gaa tgg att gat gac aaa gga gtc ttt gtt gct ggc tat gaa agc ctg				145
	Glu Trp Ile Asp Asp Lys Gly Val Phe Val Ala Gly Tyr Glu Ser Leu				
45	aaa aag aat gaa att ctt cat ctg aaa tta cct ctc aga ctt tct gta				193
	Lys Lys Asn Glu Ile Leu His Leu Lys Leu Pro Leu Arg Leu Ser Val				
	50 55 60				
	aag gaa aac aag ggc tta ttc cca gaa aga gat ttc aaa gtg cgc cat				241
50	Lys Glu Asn Lys Gly Leu Phe Pro Glu Arg Asp Phe Lys Val Arg His				
	65 70 75				
	gga gga ttt tca gac agg tct atc ttt gat cta aag cat gtg cca cat				289
	Gly Gly Phe Ser Asp Arg Ser Ile Phe Asp Leu Lys His Val Pro His				
	80 85 90				
55	acc aga ttg ctg gtt acc agt ggc ctt cca ggt tgt tat ctg cag gtg				337
	Thr Arg Leu Leu Val Thr Ser Gly Leu Pro Gly Cys Tyr Leu Gln Val				
	95 100 105 110				
	tgg cag gtt gca gag gac agt gat gtc att aaa gct gtc agc acc att				385
	Trp Gln Val Ala Glu Asp Ser Asp Val Ile Lys Ala Val Ser Thr Ile				
60	gct gtg cat gag aaa gag gag agt ctc tgg cct agg gtg gcc gtc ttc				433
	Ala Val His Glu Lys Glu Glu Ser Leu Trp Pro Arg Val Ala Val Phe				
	130 135 140				

	tcc	aca	ttg	gca	ccc	gga	gtc	ctc	cat	ggg	gcg	agg	ctc	cga	agt	ctg	481
	Ser	Thr	Leu	Ala	Pro	Gly	Val	Leu	His	Gly	Ala	Arg	Leu	Arg	Ser	Leu	
			145					150					155				
5	cag	gtc	gtt	gat	ctg	gag	tcc	cgg	aag	acc	acg	tac	acc	tca	gat	gtc	529
	Gln	Val	Val	Asp	Leu	Glu	Ser	Arg	Lys	Thr	Thr	Tyr	Thr	Ser	Asp	Val	
		160					165					170					
	agt	gac	agt	gag	gag	ctg	agt	agc	ctg	cag	gtc	cta	gat	gcg	gac	acc	577
	Ser	Asp	Ser	Glu	Glu	Leu	Ser	Ser	Leu	Gln	Val	Leu	Asp	Ala	Asp	Thr	
		175				180					185					190	
10	ttt	gcc	ttc	tgc	tgt	gct	tcg	ggc	cgg	ctg	ggg	ctt	gtt	gac	acc	cgg	625
	Phe	Ala	Phe	Cys	Cys	Ala	Ser	Gly	Arg	Leu	Gly	Leu	Val	Asp	Thr	Arg	
				195					200					205			
	cag	aag	tgg	gca	ccg	ttg	gag	aat	cgc	agc	cct	ggc	cct	ggg	tct	ggg	673
15	Gln	Lys	Trp	Ala	Pro	Leu	Glu	Asn	Arg	Ser	Pro	Gly	Pro	Gly	Ser	Gly	
				210				215					220				
	gga	gag	aga	tgg	tgt	gct	gaa	gtt	ggg	agc	tgg	ggc	cag	ggc	cct	ggg	721
	Gly	Glu	Arg	Trp	Cys	Ala	Glu	Val	Gly	Ser	Trp	Gly	Gln	Gly	Pro	Gly	
		225					230					235					
20	ccc	agc	att	gcc	agc	ctt	agc	tca	gat	ggg	cgt	ctt	tgt	ctt	ctt	gac	769
	Pro	Ser	Ile	Ala	Ser	Leu	Ser	Ser	Asp	Gly	Arg	Leu	Cys	Leu	Leu	Asp	
		240				245						250					
	ccc	cgg	gat	ctc	tgc	cat	cct	gtg	agc	tca	gtc	cag	tgc	cca	gta	tcc	817
	Pro	Arg	Asp	Leu	Cys	His	Pro	Val	Ser	Ser	Val	Gln	Cys	Pro	Val	Ser	
		255				260					265					270	
25	gta	cct	agc	cct	gac	cca	gag	ctg	ctg	cga	gtg	act	tgg	gcc	cca	ggc	865
	Val	Pro	Ser	Pro	Asp	Pro	Glu	Leu	Leu	Arg	Val	Thr	Trp	Ala	Pro	Gly	
				275						280					285		
	ctg	aag	aat	tgc	ttg	gcc	atc	tca	ggg	ttt	gat	ggg	aca	gtc	cag	gtc	913
30	Leu	Lys	Asn	Cys	Leu	Ala	Ile	Ser	Gly	Phe	Asp	Gly	Thr	Val	Gln	Val	
			290					295						300			
	tat	gat	gcc	aca	tct	tgg	gat	gga	aca	cgg	agc	caa	gat	gga	aca	cgg	961
	Tyr	Asp	Ala	Thr	Ser	Trp	Asp	Gly	Thr	Arg	Ser	Gln	Asp	Gly	Thr	Arg	
		305					310					315					
35	agc	caa	gta	gaa	cct	ctc	ttc	act	cac	aga	ggg	cac	atc	ttc	cta	gat	1009
	Ser	Gln	Val	Glu	Pro	Leu	Phe	Thr	His	Arg	Gly	His	Ile	Phe	Leu	Asp	
		320				325						330					
	gga	aat	ggg	atg	gac	cct	gct	cct	ttg	gtc	acc	acc	cac	acc	tgg	cat	1057
	Gly	Asn	Gly	Met	Asp	Pro	Ala	Pro	Leu	Val	Thr	Thr	His	Thr	Trp	His	
		335				340				345					350		
40	ccc	tgc	aga	cca	agg	act	ttg	tta	tca	gca	aca	aat	gat	gcc	tct	ctg	1105
	Pro	Cys	Arg	Pro	Arg	Thr	Leu	Leu	Ser	Ala	Thr	Asn	Asp	Ala	Ser	Leu	
				355					360					365			
	cat	gtg	tgg	gac	tgg	gtg	gac	ctt	tgt	gcc	ccc	cgc	tgacaccagc				1151
45	His	Val	Trp	Asp	Trp	Val	Asp	Leu	Cys	Ala	Pro	Arg					
		370						375									
	atctttccat	ctaggcctct	agaaaggga	ggagctgctg	tagtagcaag	gggctgctg											1211
	taggactcaa	gtgactacca	gtccctgtta	ccagctgtgt	ggccttgggc	aagtctgcca											1271
	gcgtcactta	gcctcagttt	ccttatctgt	aaaatgagga	tagtaagaac	tacctcgtag											1331
	tgatattgctg	aagggttagaa	gaaacgcatg	gcataattac	ttggtagcta	ttgttagatc											1391
50	tgggagtgtg	aaatggttagc	gttttgtccc	tgtcttcaca	ctatcatagg	gagaatcaaa											1451
	agagctaaca	aatataaaca	tgctttgtga	atTTTTTTTaa	agaaaaaaat	gtaggggggc											1511
	caataaacat	gaaaaaatcc	cagccctagt	agcaattaag	gaaatagcaa	aacaggattt											1571
	ctgctcctct	tgaggggggtc	tcattgggaac	acaggtgcac	tttcccacac	ttgtccccc											1631
	aggtagactag	gttcaagaga	catttgcttt	tggtggcccc	acaaacattt	ccttttgagg											1691
55	gcccatagtg	aatatntaaa	gtgtgctgga	catggtggct	catgcctgta	atccagcac											1751
	tttcagagtg	tgaggtgggc	agattgcttg	agctgaggag	tttgagacca	gcctgggcaa											1811
	catagcaaga	tcctttcccc	aaaaaaaaaa	aaaaaaaaa													1849
	<210>	166															
60	<211>	1748															
	<212>	DNA															
	<213>	Homo sapiens															

<220>

<221> CDS

<222> 136..264

5 <400> 166

attatttgaa aaaccactgg gttccgagtt cattactaca ggaaaaactt tctcttctgt 60
 ggcacagaga accctgcttc aaagcagaag tagcagttcc ggagtccagc tggctaaaaac 120
 tcatcccaga ggata atg gca acc cat gcc tta gaa atc gct ggg ctg ttt 171

Met Ala Thr His Ala Leu Glu Ile Ala Gly Leu Phe

10

ctt ggt ggt gtt gga atg gtg ggc aca gtg gct gtc act gtc atg cct 219
 Leu Gly Gly Val Gly Met Val Gly Thr Val Ala Val Thr Val Met Pro

15

cag tgg aga gtg tcg gcc ttc att gaa aac aac atc gtg gtt ttt 264
 Gln Trp Arg Val Ser Ala Phe Ile Glu Asn Asn Ile Val Val Phe

20

taaaactttct gggaaggact gtggatgaat tgcgtgaggc aggctaacat caggatgcag 324
 tgcaaaatct atgattccct gctggctctt tctccggacc tacaggcagc cagaggactg 384
 atgtgtgctg cttccgtgat gtccttcttg gctttcatga tggccatcct tggcatgaaa 444

25

tgcaccaggt gcacggggga caatgagaag gtgaaggctc acattctgct gacggctgga 504
 atcatcttca tcatcacggg catggtggtg ctcattccctg tgagctgggt tgccaatgcc 564
 atcatcagag atttctataa ctcaatagtg aatgttgccc aaaaacgtga gcttggagaa 624

30

gctctctact taggatggac cacggcactg gtgctgattg ttggaggagc tctgttctgc 684
 tgcgtttttt gttgcaacga aaagagcagt agctacagat actcgatacc ttcccatcgc 744
 acaaccctaaa aaagttatca caccggaaaag aagtcaccga gcgtctactc cagaagtcag 804

35

tatgtgtagt tgtgtatgtt tttttaactt tactataaag ccatgcaaat gacaaaaatc 864
 tatattactt tctcaaaatg gaccccaaaag aaactttgat ttactgttct taactgccta 924
 atcttaatta caggaactgt gcatcagcta tttatgattc tataagctat ttcagcagaa 984

40

tgagatatta aaccgaatgc tttgattgtt ctagaaagta tagtaatttg ttttctaagg 1044
 tggktcaagc atctactctt tttatcattt acttcaaaat gacattgcta aagactgcat 1104
 tatttttacta ctgtaatttc tccacgacat agcattatct acatagatga gtgtaacatt 1164

45

tatatctcac atagagacat gcttatatgg ttkcatttaa aatgaaatgc cagtccatta 1224
 cactgaataa atagaactca actattgctt ttcagggaaa tcatggatag ggttgaagaa 1284
 ggttactatt aattgtttta aaacagctta gggattaatg tcctccattt ataataaga 1344

50

ttaaaatgaa ggctttaatc agcattgtaa aggaaattga atggctttct gatatgctgt 1404
 tttttagcct aggagttaga aatcctaact tctttatcct cttctcccag aggccttttt 1464
 tttcttggtg attaaattaa cattttttaa aagcagatat tttgtcaagg ggctttgcat 1524

55

tcaaactgct tttccagggc tatactcaga agaaagataa aagtgtgatc taagaaaaag 1584
 tgatggtttt aggaaagtga aaatatTTTT gtttttTgtat ttgaagaaga atgatgcatt 1644
 ttgacaagaa atcatatatg tatggatata ttttaataaag tatttgagta cagactttga 1704

60

ggtttcatca atataaataa aagagcaaaa aaaaaaaaaa aaaa 1748

<210> 167

<211> 1275

45

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

50

<222> 14..1048

<400> 167

agaggttggg aag atg gcg tgg cga ggc tgg gcg cag aga ggc tgg ggc 49
 Met Ala Trp Arg Gly Trp Ala Gln Arg Gly Trp Gly

55

tgc ggc cag gcg tgg ggt gcg tcg gtg ggc ggc cgc agc tgc gag gag 97
 Cys Gly Gln Ala Trp Gly Ala Ser Val Gly Gly Arg Ser Cys Glu Glu

60

ctc act gcg gtc cta acc ccg ccg cag ctc ctc gga cgc agg ttt aac 145
 Leu Thr Ala Val Leu Thr Pro Pro Gln Leu Leu Gly Arg Arg Phe Asn

ttc ttt att caa caa aaa tgc gga ttc aga aaa gca ccc agg aag gtt 193
 Phe Phe Ile Gln Gln Lys Cys Gly Phe Arg Lys Ala Pro Arg Lys Val

	45				50				55				60	
	gaa cct cga aga tca	gac cca ggg aca agt ggt	gaa gca tac aag aga	241										
	Glu Pro Arg Arg	Ser Asp Pro Gly Thr	Ser Gly Glu Ala Tyr	Lys Arg										
		65			70				75					
5	agt gct ttg att cct cct gtg gaa gaa aca gtc ttt tat cct tct ccc	289												
	Ser Ala Leu Ile Pro Pro Val Glu Glu Thr Val Phe Tyr Pro Ser Pro													
		80			85				90					
	tat cct ata agg agt ctc ata aaa cct tta ttt ttt act gtt ggg ttt	337												
	Tyr Pro Ile Arg Ser Leu Ile Lys Pro Leu Phe Phe Thr Val Gly Phe													
10		95			100				105					
	aca ggc tgt gca ttt gga tca gct gct att tgg caa tat gaa tca ctg	385												
	Thr Gly Cys Ala Phe Gly Ser Ala Ala Ile Trp Gln Tyr Glu Ser Leu													
		110			115				120					
	aaa tcc agg gtc cag agt tat ttt gat ggt ata aaa gct gat tgg ttg	433												
15	Lys Ser Arg Val Gln Ser Tyr Phe Asp Gly Ile Lys Ala Asp Trp Leu													
		125			130				135					
	gat agc ata aga cca caa aaa gaa gga gac ttc aga aag gag att aac	481												
	Asp Ser Ile Arg Pro Gln Lys Glu Gly Asp Phe Arg Lys Glu Ile Asn													
		145			150				155					
20	aag tgg tgg aat aac cta agt gat ggc cag cgg act gtg aca ggt att	529												
	Lys Trp Trp Asn Asn Leu Ser Asp Gly Gln Arg Thr Val Thr Gly Ile													
		160			165				170					
	ata gct gca aat gtc ctt gta ttc tgt tta tgg aga gta cct tct ctg	577												
	Ile Ala Ala Asn Val Leu Val Phe Cys Leu Trp Arg Val Pro Ser Leu													
25		175			180				185					
	cag cgg aca atg atc aga tat ttc aca tcg aat cca gcc tca aag gtc	625												
	Gln Arg Thr Met Ile Arg Tyr Phe Thr Ser Asn Pro Ala Ser Lys Val													
		190			195				200					
	ctt tgt tct cca atg ttg ctg tca aca ttc agt cat ttc tcc tta ttt	673												
30	Leu Cys Ser Pro Met Leu Leu Ser Thr Phe Ser His Phe Ser Leu Phe													
		205			210				215					
	cac atg gca gca aat atg tat gtt ttg tgg agc ttc tct tcc agc ata	721												
	His Met Ala Ala Asn Met Tyr Val Leu Trp Ser Phe Ser Ser Ser Ile													
		225			230				235					
35	gtg aac att ctg ggt caa gag cag ttc atg gca gtg tac cta tct gca	769												
	Val Asn Ile Leu Gly Gln Glu Gln Phe Met Ala Val Tyr Leu Ser Ala													
		240			245				250					
	ggt gtt att tcc aat ttt gtc agt tac gtg ggt aaa gtt gcc aca gga	817												
	Gly Val Ile Ser Asn Phe Val Ser Tyr Val Gly Lys Val Ala Thr Gly													
40		255			260				265					
	aga tat gga cca tca ctt ggt gca gcc ctg aaa gcc att atc gcc atg	865												
	Arg Tyr Gly Pro Ser Leu Gly Ala Ala Leu Lys Ala Ile Ile Ala Met													
		270			275				280					
	gat aca gca gga atg atc ctg gga tgg aaa ttt ttt gat cat gcg gca	913												
45	Asp Thr Ala Gly Met Ile Leu Gly Trp Lys Phe Phe Asp His Ala Ala													
		285			290				295					
	cat ctt ggg gga gct ctt ttt gga ata tgg tat gtt act tac ggt cat	961												
	His Leu Gly Gly Ala Leu Phe Gly Ile Trp Tyr Val Thr Tyr Gly His													
		305			310				315					
50	gaa ctg att tgg aag aac agg gag ccg cta gtg aaa atc tgg cat gaa	1009												
	Glu Leu Ile Trp Lys Asn Arg Glu Pro Leu Val Lys Ile Trp His Glu													
		320			325				330					
	ata agg act aat ggc ccc aaa aaa gga ggt ggc tct aag taaaactggg	1058												
	Ile Arg Thr Asn Gly Pro Lys Lys Gly Gly Gly Ser Lys													
55		335			340				345					
	attggacagt agtgggtgcat ctgggtccttg ccgcctgaga gccccaggag acatcggtcta	1118												
	gagtaccat gccctatgctc ccgtctggaa gatgccagca tctggcctcc cacttttttc	1178												
	agctgtgtcc ccaggtccgt gtcttttttag aatgtgaatg atgataaagt tgtgaaataa	1238												
	aggttttctat ctagtgttgca aaaaaaaaaa aaaaaaa	1275												
60														
	<210> 168													
	<211> 1023													
	<212> DNA													

<213> Homo sapiens

<220>

<221> CDS

5 <222> 70..777

<400> 168

	aataggtccg gttccggggg cgcgtggctg cagcggggcc cgcgtggtgc ctcctgagge	60
	ggcccccg atg aag aga tct ggg aac ccg gga gcc gag gta acg aac agc	111
10	Met Lys Arg Ser Gly Asn Pro Gly Ala Glu Val Thr Asn Ser	
	1 5 10	
	tcg gtg gca ggg cct gac tgc tgc gga ggc ctc ggc aat att gat ttt	159
	Ser Val Ala Gly Pro Asp Cys Cys Gly Gly Leu Gly Asn Ile Asp Phe	
	15 20 25 30	
15	aga cag gca gac ttc tgc gtt atg acc cgg ctg ctg ggc tac gtg gac	207
	Arg Gln Ala Asp Phe Cys Val Met Thr Arg Leu Leu Gly Tyr Val Asp	
	35 40 45	
	ccc ctg gat ccc agc ttt gtg gct gcc gtc atc acc atc acc ttc aat	255
	Pro Leu Asp Pro Ser Phe Val Ala Ala Val Ile Thr Ile Thr Phe Asn	
20	50 55 60	
	ccg ctc tac tgg aat gtg gtt gca cga tgg gaa cac aag acc cgc aag	303
	Pro Leu Tyr Trp Asn Val Val Ala Arg Trp Glu His Lys Thr Arg Lys	
	65 70 75	
	ctg agc agg gcc ttc gga tcc ccc tac ctg gcc tgc tac tct cta agc	351
25	Leu Ser Arg Ala Phe Gly Ser Pro Tyr Leu Ala Cys Tyr Ser Leu Ser	
	80 85 90	
	atc acc atc ctg ctc ctg aac ttc ctg cgc tgc cac tgc ttc acg cag	399
	Ile Thr Ile Leu Leu Asn Phe Leu Arg Ser His Cys Phe Thr Gln	
	95 100 105 110	
30	gcc atg ctg agc cag ccc agg atg gag agc ctg gac acc ccc gcg gcc	447
	Ala Met Leu Ser Gln Pro Arg Met Glu Ser Leu Asp Thr Pro Ala Ala	
	115 120 125	
	tac agc ctg gtc ctc gca ctc ctg gga ctg ggc gtc gtg ctc gtg ctc	495
	Tyr Ser Leu Val Leu Ala Leu Leu Gly Leu Gly Val Val Leu Val Leu	
35	130 135 140	
	tcc agc ttc ttt gca ctg ggg ttc gct gga act ttc cta ggt gat tac	543
	Ser Ser Phe Phe Ala Leu Gly Phe Ala Gly Thr Phe Leu Gly Asp Tyr	
	145 150 155	
	ttc ggg atc ctc aag gag gcg aga gtg acc gtg ttc ccc ttc aac atc	591
40	Phe Gly Ile Leu Lys Glu Ala Arg Val Thr Val Phe Pro Phe Asn Ile	
	160 165 170	
	ctg gac aac ccc atg tac tgg gga agc aca gcc aac tac ctg ggc tgg	639
	Leu Asp Asn Pro Met Tyr Trp Gly Ser Thr Ala Asn Tyr Leu Gly Trp	
	175 180 185 190	
45	gcc atc atg cac gcc agc ccc acg ggc ctg ctc ctg acg gtg ctg gtg	687
	Ala Ile Met His Ala Ser Pro Thr Gly Leu Leu Leu Thr Val Leu Val	
	195 200 205	
	gcc ctc acc tac ata gtg gct ctc cta tac gaa gag ccc ttc acc gct	735
	Ala Leu Thr Tyr Ile Val Ala Leu Leu Tyr Glu Glu Pro Phe Thr Ala	
50	210 215 220	
	gag atc tac cgg cag aaa gcc tcc ggg tcc cac aag agg agc	777
	Glu Ile Tyr Arg Gln Lys Ala Ser Gly Ser His Lys Arg Ser	
	225 230 235	
	tgattgagct gcaacagctt tgctgaaggc ctggccagcc tcctggcctg ccccaagtgg	837
55	caggccctgc gcaggcgag aatggtgctt gctgctcagg gctcgcccc ggctggtggct	897
	gccccagtgc cttggaacct gctgccttgg ggaccctgga cgtgccgaca tatggccatt	957
	gagctccaac ccacacattc ccattcacca ataaaggcac cctgacctca aaaaaaaaaa	1017
	aaaaaa	1023
60	<210> 169	
	<211> 1085	
	<212> DNA	
	<213> Homo sapiens	

5

10 caa cag ctt ctc cat cac gcc aga aat ggc aat gct gaa gaa gta aga 103
 Gln Gln Leu Leu His His Ala Arg Asn Gly Asn Ala Glu Glu Val Arg

15 25 30 35
tgc aaa gga aga agt aag tct aac ttg ggc tgg aca ccc cta cat ctg 199

gca tgc tat ttt gga cac aga caa gtg gtc cag gat ctg ttg aag gct 247
20 Ala Cys Tyr Phe Gly His Arg Gln Val Val Gln Asp Leu Leu Lys Ala

25 cga gct gcc ttt ⁷⁵aca gga cga aag gtg ⁸⁰aaa atc att cta tgt ⁸⁵tca atg 343

ttt gta agt gag gta ttt gga gga gta gtt acc att gtt ttc tct gtt 391
Phe Val Ser Glu Val Phe Gly Gly Val Val Thr Ile Val Phe Ser Val

Ile Thr Ile
120

tgtgtactcc	agcaatgggg	aactcctgat	tattcttttt	ttttaaaaaa	aaatagcttc	620
attgaggtat	aacttacatt	gcataaacct	cacctgtgat	atttgtgaaa	atatatttgg	680
tttttttttt	tctctctctt	cctatctctt	ccagcttgat	ggccctgaat	gcatctcagg	740

40	tttgcAACCT	ccaggggaggG	tagagggggct	gaagggggaaa	tggctcgcta	atggccagtG	880
	gttttcatcaa	tcatgcctat	ttaatggaac	ctccataaaa	acctgaaagg	acagggttct	920
	ccagagctact	cggtgcctca	acaggtggag	gttcttgaat	gcatcacacgg	acagggggca	980

45 gtaataaagc agtaaacatg tttttcttgaa aaaaaaaaaa aaaaaa 1005
(210) 170

50 <213> Homo sapiens

<222> 63..572

ga atg cgg ctt caa ggg gct atc ttt gtg ctc ctg ccc cac ctg ggg 107
Met Arg Leu Gln Gly Ala Ile Phe Val Leu Leu Pro His Leu Gly

Pro	Ile	Leu	Val	Trp	Leu	Phe	Thr	Arg	Asp	His	Met	Ser	Gly	Trp	Cys
				20					25					30	

177

178

	90	95	100	
	aac tta ctg gca agc tcg gac gct gcc ctt tca gcc tcc atg gcc agc			510
	Asn Leu Leu Ala Ser Ser Asp Ala Ala Leu Ser Ala Ser Met Ala Ser			
	105	110	115	
5	ctc ctg gag gac ctc agc cac att gag ggc ctg agt cag gct ccc caa			558
	Leu Leu Glu Asp Leu Ser His Ile Glu Gly Leu Ser Gln Ala Pro Gln			
	120	125	130	
	ccc ttg gca gac gag ggg cca cca ggc cgt agc atc ggg gga gca gcg			606
	Pro Leu Ala Asp Glu Gly Pro Pro Gly Arg Ser Ile Gly Gly Ala Ala			
10	135	140	145	
	ccc agc ctg ggt gcc ttg gac ctg ctg ggc cca gcc act ggc tgt cta			654
	Pro Ser Leu Gly Ala Leu Asp Leu Leu Gly Pro Ala Thr Gly Cys Leu			
	150	155	160	165
	ctg gac gat ggg ctt gag ggc ctg ttt gag gat att gac acc tct atg			702
15	Leu Asp Asp Gly Leu Glu Gly Leu Phe Glu Asp Ile Asp Thr Ser Met			
	170	175	180	
	tat gac aat gaa ctt tgg gca cca gcc tct gag ggc ctc aaa cca ggc			750
	Tyr Asp Asn Glu Leu Trp Ala Pro Ala Ser Glu Gly Leu Lys Pro Gly			
	185	190	195	
20	cct gag gat ggg ccg ggc aag gag gaa gct ccg gag ctg gac gag gcc			798
	Pro Glu Asp Gly Pro Gly Lys Glu Glu Ala Pro Glu Leu Asp Glu Ala			
	200	205	210	
	gaa ttg gac tac ctc atg gat gtg ctg gtg ggc aca cag gca ctg gag			846
	Glu Leu Asp Tyr Leu Met Asp Val Leu Val Gly Thr Gln Ala Leu Glu			
25	215	220	225	
	cga ccg ccg ggg cca ggg cgc tgagccctcg tgctggaatg gttgtctggt			897
	Arg Pro Pro Gly Pro Gly Arg			
	230	235		
	atctgaactg agcctgctgg ctggaccaac tgtcctcgaa aagacacagc tggcttcct			957
30	agtacagaga acagggcttg ggccactttg gagagacaga atctagtcct gggcaacttc			1017
	acatccgtcc tctgtctca gggctggcag ggggagcctg gaattacccc ctagtgatgg			1077
	aatgacaggg tctggtgggg acttaattcc ctggccctgg ggtcatagct tgggctgttc			1137
	cttctctgat acgggaagag accccaatca gatttttcaa attaaagcca gtcctgggaa			1197
	atctcaaaaa aaaaaaaaaa aa			1219
35	<210> 172			
	<211> 1487			
	<212> DNA			
	<213> Homo sapiens			
40	<220>			
	<221> CDS			
	<222> 68..640			
45	<400> 172			
	gacgaaggac tggaagggtgg cggtggtgaa ggtgcaggcc gttggggcgg ctcagaggca			60
	ggtgact atg aaa ggc tta tat ttc caa cag agt tcc aca gat gaa gaa			109
	Met Lys Gly Leu Tyr Phe Gln Gln Ser Ser Thr Asp Glu Glu			
	1	5	10	
50	ata aca ttt gta ttt caa gaa aag gaa gat ctt cct gtt aca gag gat			157
	Ile Thr Phe Val Phe Gln Glu Lys Glu Asp Leu Pro Val Thr Glu Asp			
	15	20	25	30
	aac ttt gtg aaa ctt caa gtt aaa gct tgt gct ctg agc cag ata aat			205
	Asn Phe Val Lys Leu Gln Val Lys Ala Cys Ala Leu Ser Gln Ile Asn			
55	35	40	45	
	aca aag ctt ctg gca gaa atg aag atg aaa aag gat tta ttt cct gtt			253
	Thr Lys Leu Leu Ala Glu Met Lys Met Lys Lys Asp Leu Phe Pro Val			
	50	55	60	
	ggg aga gaa att gct gga att gta tta gat gtt gga agc aag gta tca			301
60	Gly Arg Glu Ile Ala Gly Ile Val Leu Asp Val Gly Ser Lys Val Ser			
	65	70	75	
	ttc ttt caa cca gat gat gaa gta gtt gga att ttg ccc ctg gac tct			349
	Phe Phe Gln Pro Asp Asp Glu Val Val Gly Ile Leu Pro Leu Asp Ser			

180

	aga gtc atc tgt ttg gtg gga gct gga atc tcc aca tcc gca ggc atc	410
	Arg Val Ile Cys Leu Val Gly Ala Gly Ile Ser Thr Ser Ala Gly Ile	
	80 85 90	
5	ccc gac ttt cgc tct cca tcc acc ggc ctc tat gac aac cta gag aag	458
	Pro Asp Phe Arg Ser Pro Ser Thr Gly Leu Tyr Asp Asn Leu Glu Lys	
	95 100 105	
	tac cat ctt ccc tac cca gag gcc atc ttt gag atc agc tat ttc aag	506
	Tyr His Leu Pro Tyr Pro Glu Ala Ile Phe Glu Ile Ser Tyr Phe Lys	
	110 115 120 125	
10	aaa cat ccg gaa ccc ttc ttc gcc ctc gcc aag gaa ctc tat cct ggg	554
	Lys His Pro Glu Pro Phe Phe Ala Leu Ala Lys Glu Leu Tyr Pro Gly	
	130 135 140	
	cag ttc aag cca acc atc tgt cac tac ttc atg cgc ctg ctg aag gac	602
	Gln Phe Lys Pro Thr Ile Cys His Tyr Phe Met Arg Leu Leu Lys Asp	
	145 150 155	
15	aag ggg cta ctc ctg cgc tgc tac acg cag aac ata gat acc ctg gag	650
	Lys Gly Leu Leu Leu Arg Cys Tyr Thr Gln Asn Ile Asp Thr Leu Glu	
	160 165 170	
20	cga ata gcc ggg ctg gaa cag gag gac ttg gtg gag gcg cac ggc acc	698
	Arg Ile Ala Gly Leu Glu Gln Glu Asp Leu Val Glu Ala His Gly Thr	
	175 180 185	
	ttc tac aca tca cac tgc gtc agc gcc agc tgc cgg cac gaa tac ccg	746
	Phe Tyr Thr Ser His Cys Val Ser Ala Ser Cys Arg His Glu Tyr Pro	
	190 195 200 205	
25	cta agc tgg atg aaa gag aag atc ttc tct gag gtg acg ccc aag tgt	794
	Leu Ser Trp Met Lys Glu Lys Ile Phe Ser Glu Val Thr Pro Lys Cys	
	210 215 220	
	gaa gac tgt cag agc ctg gtg aag cct gat atc gtc ttt ttt ggt gag	842
	Glu Asp Cys Gln Ser Leu Val Lys Pro Asp Ile Val Phe Phe Gly Glu	
	225 230 235	
30	agc ctc cca gcg cgt ttc ttc tcc tgt atg cag tca gac ttc ctg aag	890
	Ser Leu Pro Ala Arg Phe Phe Ser Cys Met Gln Ser Asp Phe Leu Lys	
	240 245 250	
	gtg gac ctc ctc ctg gtc atg ggt acc tcc ttg cag gtg cag ccc ttt	938
	Val Asp Leu Leu Leu Val Met Gly Thr Ser Leu Gln Val Gln Pro Phe	
	255 260 265	
	gcc tcc ctc atc agc aag gca ccc ctc tcc acc cct cgc ctg ctc atc	986
	Ala Ser Leu Ile Ser Lys Ala Pro Leu Ser Thr Pro Arg Leu Leu Ile	
	270 275 280 285	
40	aac aag gag aaa gct ggc cag tcg gac cct ttc ctg ggg atg att atg	1034
	Asn Lys Glu Lys Ala Gly Gln Ser Asp Pro Phe Leu Gly Met Ile Met	
	290 295 300	
	ggc ctc gga gga ggc atg gac ttt gac tcc aag aag gcc tac agg gac	1082
	Gly Leu Gly Gly Gly Met Asp Phe Asp Ser Lys Lys Ala Tyr Arg Asp	
	305 310 315	
45	gtg gcc tgg ctg ggt gaa tgc gac cag ggc tgc ctg gcc ctt gct gag	1130
	Val Ala Trp Leu Gly Glu Cys Asp Gln Gly Cys Leu Ala Leu Ala Glu	
	320 325 330	
	ctc ctt gga tgg aag aag gag ctg gag gac ctt gtc cgg agg gag cac	1178
	Leu Leu Gly Trp Lys Lys Glu Leu Glu Asp Leu Val Arg Arg Glu His	
	335 340 345	
	gcc agc ata gat gcc cag tcg ggg gcg ggg gtc ccc aac ccc agc act	1226
	Ala Ser Ile Asp Ala Gln Ser Gly Ala Gly Val Pro Asn Pro Ser Thr	
	350 355 360 365	
55	tca gct tcc ccc aag aag tcc ccg cca cct gcc aag gac gag gcc agg	1274
	Ser Ala Ser Pro Lys Lys Ser Pro Pro Pro Ala Lys Asp Glu Ala Arg	
	370 375 380	
	aca aca gag agg gag aaa ccc cag tgacagctgc atctcccagg cgggatgccg	1328
	Thr Thr Glu Arg Glu Lys Pro Gln	
	385	
60	agctcctcag ggacagctga gcccacaaccg ggccctggccc cctcttaacc agcagttctt	1388
	gtctggggag ctccagaacat cccccaatct ctacagctc cctccccaac actgggggtcc	1448
	cagcaaccct ggcccccaac cccagcaaat ctctaaccac tcctagaggc caaggcttaa	1508

	acaggcatct	ctaccagccc	cactgtctct	aaccactcct	gggctaagga	gtaacctccc	1568
	tcattctctaa	ctgccccccac	ggggccagg	ctaccccaga	acttttaact	cttccaggac	1628
	agggagcttc	gggccccccac	tctgtctcct	gcccccg	gcctgtggct	aagtaaacca	1688
	tacctaacct	acccagtg	gggtgtgggc	ctctgaatct	aaccacacc	cagcgtaggg	1748
5	ggagtctgag	ccgggagggc	tcccgagtct	ctgccttcag	ctcccaaagt	gggtggtggg	1808
	cccccttcac	gtgggaccca	cttcccatgc	tggatgggca	gaagacattg	cttattggag	1868
	acaaattaaa	aacaaaaaca	actaacaag	aaaaaaaaa	aaaaaaa		1915
	<210> 174						
10	<211> 1990						
	<212> DNA						
	<213> Homo sapiens						
	<220>						
15	<221> CDS						
	<222> 259..1701						
	<400> 174						
20	ttagaacatc	yttaatcaaaa	aatttttggtg	cgagagaaac	aataggacgg	aaacgcccag	60
	gaacccggct	gaggcgccag	cttcctaggt	gacagacagg	tacactgtat	gctagccctg	120
	tatctgtctg	agcagtggaa	tgtgccagga	aagaaggagc	aaccactgac	tgatgaacct	180
	ttgcagctct	cccttccaag	agggatgcca	gagccttctg	taagctcctc	agatgtcact	240
	ggatatctagg	caacaggg	atg agc ctg	aac ctc cct	gag gcc agc	tta ctt	291
			Met Ser Leu Asn Leu Pro	Glu Ala Ser Leu Leu			
25			1	5	10		
	agc aga gca tcc tgg cca gaa caa gcc aag gag cca aga cga gag gga						339
	Ser Arg Ala Ser Trp Pro Glu Gln Ala Lys Glu Pro Arg Arg Glu Gly						
		15	20	25			
30	cac acg gac aaa caa cag aca gaa gac gta ctg gcc gct gga ctc cgc						387
	His Thr Asp Lys Gln Gln Thr Glu Asp Val Leu Ala Ala Gly Leu Arg						
		30	35	40			
	tgc ctc ccc cat ctc ccc gcc atc tgc gcc cgg agg atg agc cca gcc						435
	Cys Leu Pro His Leu Pro Ala Ile Cys Ala Arg Arg Met Ser Pro Ala						
		45	50	55			
35	ttc agg gcc atg gat gtg gag ccc cgc gca aaa ggc gtc ctt ctg gag						483
	Phe Arg Ala Met Asp Val Glu Pro Arg Ala Lys Gly Val Leu Leu Glu						
		60	65	70	75		
	ccc ttt gtc cac cag gtc ggg ggg cac tca tgc gtg ctc cgc ttc aat						531
	Pro Phe Val His Gln Val Gly Gly His Ser Cys Val Leu Arg Phe Asn						
40		80	85	90			
	gag aca acc ctg tgc aag ccc ctg gtc cca agg gaa cat cag ttc tac						579
	Glu Thr Thr Leu Cys Lys Pro Leu Val Pro Arg Glu His Gln Phe Tyr						
		95	100	105			
45	gag acc ctc cct gct gag atg cgc aaa ttc act ccc cag tac aaa ggt						627
	Glu Thr Leu Pro Ala Glu Met Arg Lys Phe Thr Pro Gln Tyr Lys Gly						
		110	115	120			
	gtg gta tct gtg cgc ttt gaa gaa gat gaa gac agg aac ttg tgt cta						675
	Val Val Ser Val Arg Phe Glu Glu Asp Glu Asp Arg Asn Leu Cys Leu						
		125	130	135			
50	ata gca tat cca ttg aaa ggg gac cat gga att gtg gac att gta gat						723
	Ile Ala Tyr Pro Leu Lys Gly Asp His Gly Ile Val Asp Ile Val Asp						
		140	145	150	155		
	aat tca gac tgt gaa cca aaa agt aag ctc cta agg tgg aca aca aac						771
	Asn Ser Asp Cys Glu Pro Lys Ser Lys Leu Arg Trp Thr Thr Asn						
55		160	165	170			
	aaa aaa cat cat gtc tta gaa aca gaa aag acc cct aag gac tgg gtg						819
	Lys Lys His His Val Leu Glu Thr Glu Lys Thr Pro Lys Asp Trp Val						
		175	180	185			
	cgt cag cac cgt aaa gag gag aaa atg aag agc cat aag tta gaa gaa						867
60	Arg Gln His Arg Lys Glu Glu Lys Met Lys Ser His Lys Leu Glu Glu						
		190	195	200			
	gaa ttt gag tgg cta aag aaa tct gaa gtc ttg tac tac act gta gag						915
	Glu Phe Glu Trp Leu Lys Lys Ser Glu Val Leu Tyr Tyr Thr Val Glu						

	205				210				215								
	aag	aag	ggg	aat	ata	agt	tcc	cag	ctt	aaa	cac	tat	aac	cct	tgg	agc	963
	Lys	Lys	Gly	Asn	Ile	Ser	Ser	Gln	Leu	Lys	His	Tyr	Asn	Pro	Trp	Ser	
	220					225					230					235	
5	atg	aaa	tgt	cac	cag	caa	cag	tta	cag	aga	atg	aag	gag	aat	gca	aag	1011
	Met	Lys	Cys	His	Gln	Gln	Gln	Leu	Gln	Arg	Met	Lys	Glu	Asn	Ala	Lys	
					240					245					250		
	cat	cgg	aac	cag	tac	aaa	ttt	atc	tta	ctg	gaa	aac	ctg	act	tcc	cgc	1059
	His	Arg	Asn	Gln	Tyr	Lys	Phe	Ile	Leu	Leu	Glu	Asn	Leu	Thr	Ser	Arg	
10					255				260					265			
	tat	gag	gtg	cct	tgt	gtc	ctt	gac	ctc	aag	atg	ggc	aca	cga	caa	cat	1107
	Tyr	Glu	Val	Pro	Cys	Val	Leu	Asp	Leu	Lys	Met	Gly	Thr	Arg	Gln	His	
					270				275				280				
	ggg	gat	gat	gct	tca	gag	gag	aag	gca	gcc	aac	cag	atc	cga	aaa	tgt	1155
15	Gly	Asp	Asp	Ala	Ser	Glu	Glu	Lys	Ala	Ala	Asn	Gln	Ile	Arg	Lys	Cys	
												295					
	cag	cag	agc	aca	tct	gca	gtc	att	ggg	gtg	cgt	gtg	tgt	ggc	atg	cag	1203
	Gln	Gln	Ser	Thr	Ser	Ala	Val	Ile	Gly	Val	Arg	Val	Cys	Gly	Met	Gln	
												310				315	
20	gtg	tac	caa	gca	ggc	agt	ggg	cag	ctc	atg	ttc	atg	aac	aag	tac	cat	1251
	Val	Tyr	Gln	Ala	Gly	Ser	Gly	Gln	Leu	Met	Phe	Met	Asn	Lys	Tyr	His	
												325				330	
	gga	cgg	aag	cta	tgc	gtg	cag	ggc	ttc	aag	gag	gca	ctt	ttc	cag	ttc	1299
	Gly	Arg	Lys	Leu	Ser	Val	Gln	Gly	Phe	Lys	Glu	Ala	Leu	Phe	Gln	Phe	
25																	
	ttc	cac	aat	ggg	cgg	tac	ctg	cgc	cgt	gaa	ctc	ctg	ggc	cct	gtg	ctc	1347
	Phe	His	Asn	Gly	Arg	Tyr	Leu	Arg	Arg	Glu	Leu	Leu	Gly	Pro	Val	Leu	
	aag	aag	ctg	act	gag	ctc	aag	gca	gtg	ttg	gag	cga	cag	gag	tcc	tac	1395
30	Lys	Lys	Leu	Thr	Glu	Leu	Lys	Ala	Val	Leu	Glu	Arg	Gln	Glu	Ser	Tyr	
	cgc	ttc	tac	tca	agc	tcc	ctg	ctg	gtc	att	tat	gat	ggc	aag	gag	cgg	1443
	Arg	Phe	Tyr	Ser	Ser	Ser	Leu	Leu	Val	Ile	Tyr	Asp	Gly	Lys	Glu	Arg	
35	ccc	gaa	gtg	gtc	ctg	gac	tca	gat	gct	gag	gat	ttg	gag	gac	ctg	tca	1491
	Pro	Glu	Val	Val	Leu	Asp	Ser	Asp	Ala	Glu	Asp	Leu	Glu	Asp	Leu	Ser	
	gag	gaa	tca	gct	gat	gag	tct	gct	ggg	gcc	tat	gcc	tac	aaa	ccc	atc	1539
40	Glu	Glu	Ser	Ala	Asp	Glu</											

	Gly	Gly	His	Ile	Val	Ala	Ile	Gly	Gly	Leu	Gly	Asn	Gln	Pro	Cys	Pro	
	280					285				290					295		
	ttg	ggc	tct	gtg	gag	agc	ttt	agc	ctt	gca	cgg	cgg	cgc	tgg	gag	gca	1145
	Leu	Gly	Ser	Val	Glu	Ser	Phe	Ser	Leu	Ala	Arg	Arg	Arg	Trp	Glu	Ala	
5					300					305					310		
	ttg	cct	gcc	atg	ccc	act	gcc	cgc	tgc	tcc	tgc	tct	agt	ctg	cag	gct	1193
	Leu	Pro	Ala	Met	Pro	Thr	Ala	Arg	Cys	Ser	Cys	Ser	Ser	Leu	Gln	Ala	
					315					320					325		
	ggg	ccc	cgg	ctg	ttt	gtt	att	ggg	ggt	gtg	gcc	cag	ggc	ccc	agt	caa	1241
10	Gly	Pro	Arg	Leu	Phe	Val	Ile	Gly	Gly	Val	Ala	Gln	Gly	Pro	Ser	Gln	
			330					335					340				
	gcc	gtg	gag	gca	ctg	tgt	ctg	cgt	gat	ggg	gtc	tgaaggcttg	gtgggagctg				1294
	Ala	Val	Glu	Ala	Leu	Cys	Leu	Arg	Asp	Gly	Val						
					345			350									
15	tccactggag	cagctcattg	ccagaggcag	ctatttctat	ggctcctttt	gctgctgagg											1354
	acactcactg	tggctctgtg	ggatgagaga	ggcatggggg	tgagcacttg	aaactactgcc											1414
	ttggggcctt	gggttagggg	agcctttgtc	tttagtgtag	gacacacata	tgcttacacc											1474
	tacctttatc	accattcggt	catgaatcat	gcctagctcc	atccttgccc	tgggacctac											1534
	taggccttcc	atccaactgg	gaaatgggga	gaagcaaagc	tggcctcatg	ctcttcaggg											1594
20	tcagttccta	tctggagttg	accaggccta	ccccagttgc	cattcctgaa	aaatctcagc											1654
	tgccaggctg	cctttagggt	ccctgtagac	ccaggagagt	tgagagggtg	ggggacacag											1714
	agagaataga	gaggatgtgg	gaactgccag	agggccggag	cgcaggagtt	caagtggagg											1774
	aatgctggct	ttgagccctc	tacactgctg	gttgtatgac	cttggaacaag	tcacttcacc											1834
	tctctgtgcc	tcagcatcct	catctataaa	tggggatctc	tgaaaccttc	ctaccctacc											1894
25	tacctcacag	ggctgttgtg	aggacccagg	gagtttggat	gtggaagtaa	aagtgtgcc											1954
	aaaaaaaaa	aaaaaaa															1971
	<210> 176																
	<211> 1613																
30	<212> DNA																
	<213> Homo sapiens																
	<220>																
	<221> CDS																
35	<222> 68..127																
	<400> 176																
	gacgaaggac	tggaagggtg	cggtggtgaa	ggtgcaggcc	gttggggcgg	ctcagaggca											60
	ggtgact	atg	aaa	ggc	tta	tat	ttc	caa	cag	agt	tcc	aca	gat	gaa	gaa		109
40		Met	Lys	Gly	Leu	Tyr	Phe	Gln	Gln	Ser	Ser	Thr	Asp	Glu	Glu		
		1				5					10						
	ata	aca	ttt	gta	ttt	caa	taaaaggaag	atcttctctgt	tacagaggat								157
	Ile	Thr	Phe	Val	Phe	Gln											
	15					20											
45	aactttgtga	aacttcaagt	taaagcttgt	gctctgagcc	agataaatac	aaagcttctg											217
	gcagaaatga	agatgaaaaa	ggatttattt	cctgttggga	gagaaattgc	tggaattgta											277
	ttagatgttg	gaagcaagg	atcattcttt	caaccagatg	atgaagtagt	tggaattttg											337
	cccctggact	ctgaagaccc	tggactttgt	gaagttgtta	gagtacatga	gcattacttg											397
	gttcataaac	cagaaaagg	cacatggacg	gaagcagcag	gaagcattcg	ggatggagtg											457
50	cgtgcctata	cagctctgca	ttatctttct	catctctctc	ctggaaaatc	agtgtctgata											517
	atggatggag	caagtgcatt	tggtacaata	gctattcagt	tagcacatca	tagaggagcc											577
	aaagtgattt	caacagcatg	cagccttgaa	gataagcagt	gccttgaaag	attcagacct											637
	cccatagccc	gagtgattga	tgtatcta	gggaaagtgc	atggtgctga	aagctgtttg											697
	gaagaaacag	gtggcctggg	agtagatat	gtcctagatg	ctggagttag	attatatagt											757
55	aaagatgatg	aaccagctgt	aaaactacaa	ctactaccac	ataaacatga	tatcatcaca											817
	cttcttggtg	ctggaggcca	ctgggtaaca	acagaagaaa	accttcagtt	ggatcctcca											877
	gatagccat	gccttttctt	caaggagca	acgttagctt	tcctgaatga	tgaagtttgg											937
	aatttgtcaa	atgtacaaca	gggaamaata	tctttgtatc	ttaaaggatg	tgatggagaa											997
	gttatcaact	ggtgttttca	gacctcagtt	ggatgaaccc	attccactgt	atgaggcaaa											1057
60	agtttccatg	gaagctgttc	agaaaaatca	aggaagaaaa	aagcaagttg	ttcaatttta											1117
	attttcttct	ttctcagacc	tcagtcggat	gaacatatcc	cagtatttga	agccagaatt											1177
	ttctttggaa	attgttgaga	aaaaccaagg	aagataaaac	aagttgcatt	tttaagcacg											1237
	tttctctgct	aagacaagat	gctcagttga	cacatttgaa	aagtgtttga	aaaattcttg											1297

tgcaaatgat caagataatt ctataattaa catcttaagg gaatttttct aaaaaccttt 1357
 tcattgtttc tatatatattt gcccttgcta taaaattcct tccatgaaga aaactgctgc 1417
 tttcagcaaa agtcacacta ctcttgataa aagctgttgc aggcctttgc taagctatca 1477
 aagtaacgta ttaattttgt atcaactccg ttctcaacac cttccttaag tctttgctgt 1537
 5 cataatttaa gcatttgagt atattttgaa gtcttaaaag acttagccca taggcactca 1597
 aaaaaaaaaa aaaaaa 1613

 <210> 177
 <211> 1361
 10 <212> DNA
 <213> Homo sapiens

 <220>
 <221> CDS
 15 <222> 65..1024

 <400> 177
 gaaggactgg aaggtggcgg tgggtgaagg gcaggccggt ggggcggctc agaggcaggt 60
 gact atg aaa ggc tta tat ttc caa cag agt tcc aca gat gaa gaa ata 109
 20 Met Lys Gly Leu Tyr Phe Gln Gln Ser Ser Thr Asp Glu Glu Ile
 1 5 10 15
 aca ttt gta ttt caa gaa aag gaa gat ctt cct gtt aca gag gat aac 157
 Thr Phe Val Phe Gln Glu Lys Glu Asp Leu Pro Val Thr Glu Asp Asn
 20 25 30
 25 ttt gtg aaa ctt caa gtt aaa gct tgt gct ctg agc cag ata aat aca 205
 Phe Val Lys Leu Gln Val Lys Ala Cys Ala Leu Ser Gln Ile Asn Thr
 35 40 45
 aag ctt ctg gca gaa atg aag atg aaa aag gat tta ttt cct gtt ggg 253
 Lys Leu Leu Ala Glu Met Lys Met Lys Lys Asp Leu Phe Pro Val Gly
 50 55 60
 30 aga gaa att gct gga att gta tta gat gtt gga agc aag gta tca ttc 301
 Arg Glu Ile Ala Gly Ile Val Leu Asp Val Gly Ser Lys Val Ser Phe
 65 70 75
 ttt caa cca gat gat gaa gta gtt gga att ttg ccc ctg gac tct gaa 349
 35 Phe Gln Pro Asp Asp Glu Val Val Gly Ile Leu Pro Leu Asp Ser Glu
 80 85 90 95
 gac cct gga ctt tgt gaa gtt gtt aga gta cat gag cat tac ttg gtt 397
 Asp Pro Gly Leu Cys Glu Val Val Arg Val His Glu His Tyr Leu Val
 100 105 110
 40 cat aaa cca gaa aag gtc aca tgg acg gaa gca gca gga agc att cgg 445
 His Lys Pro Glu Lys Val Thr Trp Thr Glu Ala Ala Gly Ser Ile Arg
 115 120 125
 gat gga gtg cgt gcc tat aca gct ctg cat tat ctt tct cat ctc tct 493
 45 Asp Gly Val Arg Ala Tyr Thr Ala Leu His Tyr Leu Ser His Leu Ser
 130 135 140
 cct gga aaa tca gtg ctg ata atg gat gga gca agt gca ttt ggt aca 541
 Pro Gly Lys Ser Val Leu Ile Met Asp Gly Ala Ser Ala Phe Gly Thr
 145 150 155
 ata gct att cag tta gca cat cat aga gga gcc aaa gtg att tca aca 589
 50 Ile Ala Ile Gln Leu Ala His His Arg Gly Ala Lys Val Ile Ser Thr
 160 165 170 175
 gca tgc agc ctt gaa gat aag cag tgc ctt gaa aga ttc aga cct ccc 637
 Ala Cys Ser Leu Glu Asp Lys Gln Cys Leu Glu Arg Phe Arg Pro Pro
 180 185 190
 55 ata gcc cga gtg att gat gta tct aat ggg aaa gtt cat gtt gct gaa 685
 Ile Ala Arg Val Ile Asp Val Ser Asn Gly Lys Val His Val Ala Glu
 195 200 205
 agc tgt ttg gaa gaa aca ggt ggc ctg gga gta gat att gtc cta gat 733
 Ser Cys Leu Glu Glu Thr Gly Gly Leu Gly Val Asp Ile Val Leu Asp
 210 215 220
 60 gct gga gtg aga tta tat agt aaa gat gat gaa cca gct gta aaa cta 781
 Ala Gly Val Arg Leu Tyr Ser Lys Asp Asp Glu Pro Ala Val Lys Leu
 225 230 235

	caa cta cta cca cat aaa cat gat atc atc aca ctt ctt ggt gtt gga	829
	Gln Leu Leu Pro His Lys His Asp Ile Ile Thr Leu Leu Gly Val Gly	
	240 245 250 255	
5	ggc cac tgg gta aca aca gaa gaa aac ctt cag ttg gat cct cca gat	877
	Gly His Trp Val Thr Thr Glu Glu Asn Leu Gln Leu Asp Pro Pro Asp	
	260 265 270	
	agc cac tgc ctt ttc ctc aag gga gca acg tta gct ttc ctg aat gat	925
	Ser His Cys Leu Phe Leu Lys Gly Ala Thr Leu Ala Phe Leu Asn Asp	
	275 280 285	
10	gaa gtt tgg aat ttg tca aat gta caa cag gga aaa tat ctt tat ctt	973
	Glu Val Trp Asn Leu Ser Asn Val Gln Gln Gly Lys Tyr Leu Tyr Leu	
	290 295 300	
	aaa gga tgt gat gga gaa gtt atc aac tgg tgt ttt cag acc tca gtc	1021
	Lys Gly Cys Asp Gly Glu Val Ile Asn Trp Cys Phe Gln Thr Ser Val	
15	305 310 315	
	gga tgaacatatt ccagtatttg aagccagaat tttctttgga aattgttgag	1074
	Gly	
	320	
20	aaaaaccaag gaagataaaa caagttgcat ttttaagcac gtttctctgc taagacaaga	1134
	tgctcagttg acacatttga aaagtgtttg aaaaattctg gcttctaatac ctgcctctgt	1194
	tcccttttct ctccttgaaa gtccagcaca ccattcttgt ccttccccag tttctctgcc	1254
	ctccaccct ccagcttcat gctcagtggt gtgcttaata aaatggacat atttttctct	1314
	aaaaaaaaa aaaaaakaaa aaaaaaaaaat aaaaaaaaaa aaaaaaa	1361
25	<210> 178	
	<211> 1113	
	<212> DNA	
	<213> Homo sapiens	
30	<220>	
	<221> CDS	
	<222> 109..585	
	<400> 178	
35	gcgggaccgg acttccggct ggtctgtggg gtttcggggtt cgggggtttcc tgggtgggcgt	60
	caggggcagg caacagagtg gcggccgcta cggccctgga acgggggcc atg gag aag	117
	Met Glu Lys	
	1	
40	ctg cgg cga gtc ctg agc ggc cag gac gac gag gag cag ggc ctg act	165
	Leu Arg Arg Val Leu Ser Gly Gln Asp Asp Glu Glu Gln Gly Leu Thr	
	5 10 15	
	gcg cag gtc ctg gat gcc tca tcc ctt agt ttc aac acc aga ttg aaa	213
	Ala Gln Val Leu Asp Ala Ser Ser Leu Ser Phe Asn Thr Arg Leu Lys	
	20 25 30 35	
45	tgg ttt gcc atc tgc ttc gta tgt ggc gtt ttc ttt tct att ctt gga	261
	Trp Phe Ala Ile Cys Phe Val Cys Gly Val Phe Phe Ser Ile Leu Gly	
	40 45 50	
	act gga ttg ctg tgg ctt ccg ggc ggc ata aag ctt ttt gca gtg ttt	309
	Thr Gly Leu Leu Trp Leu Pro Gly Gly Ile Lys Leu Phe Ala Val Phe	
50	55 60 65	
	tat acc ctc ggc aat ctt gct gcg tta gcc agt aca tgc ttt tta atg	357
	Tyr Thr Leu Gly Asn Leu Ala Ala Leu Ala Ser Thr Cys Phe Leu Met	
	70 75 80	
	gga cct gtg aag caa ctg aag aaa atg ttt gaa gca aca aga ttg ctt	405
55	Gly Pro Val Lys Gln Leu Lys Lys Met Phe Glu Ala Thr Arg Leu Leu	
	85 90 95	
	gca aca att gtt atg ctt ttg tgt ttc ata ttt acc ctg tgt gct gct	453
	Ala Thr Ile Val Met Leu Leu Cys Phe Ile Phe Thr Leu Cys Ala Ala	
	100 105 110 115	
60	ctt tgg tgg cat aag aag gga ctg gct gtg tta ttc tgc ata ttg cag	501
	Leu Trp Trp His Lys Lys Gly Leu Ala Val Leu Phe Cys Ile Leu Gln	
	120 125 130	
	ttc ttg tca atg acc tgg tat agc ctg tcg tac atc cca tat gca agg	549

Phe Leu Ser Met Thr Trp Tyr Ser Leu Ser Tyr Ile Pro Tyr Ala Arg
 135 140 145
 gat gca gtt att aaa tgc tgt tct tct ctc cta agt tgaaaaatcag 595
 Asp Ala Val Ile Lys Cys Cys Ser Ser Leu Leu Ser
 5 150 155
 aaacattgtg gaaaagagca cttgaatgta tgggtactcta tgttttggtga agtttgcttt 655
 tccccataaa acactccagg aacaactgac gtgacagttg aagaccgttt tgtactaagt 715
 ctcatTTTTgt atactggtaa aaactacatg cttgattaaa ccattaaatg cttgtaactt 775
 taaattcatt atgtgtcatt aatatacttt tccaaagata agattttttaa tcaactgccag 835
 10 ttgtaaatta ttttttagcca attttttaaat ctttttcaaag cagcttttgaa atgtgaatat 895
 ttaaaggtag acctcgtgct gcaagataat taaacttttt tgctttttaa aaatgtctgc 955
 atttttaaga ttttttttac tttaaatgtg aaactttatt taagctagaa amattgctta 1015
 ttatatgtaa taaaaataat atataaatct ttacaatktt tgaaataaac ccaccttggt 1075
 aaaaataaaa aaaaaaaaaa agaaaaaaaa aaaaaaaa 1113
 15
 <210> 179
 <211> 1960
 <212> DNA
 <213> Homo sapiens
 20
 <220>
 <221> CDS
 <222> 29..577
 25 <400> 179
 atcggccaac ggacgcgagg cgcgcgcc atg gaa cag cgg tta gct gag ttt 52
 Met Glu Gln Arg Leu Ala Glu Phe
 1 5
 30 cgg gcg gcg cgg aaa cgg gcg ggt ctg gcg gcc caa ccc cct gct gcc 100
 Arg Ala Ala Arg Lys Arg Ala Gly Leu Ala Ala Gln Pro Pro Ala Ala
 10 15 20
 agt cag ggc gca caa acc cca gga gag aag gcg gaa gca gca gcg act 148
 Ser Gln Gly Ala Gln Thr Pro Gly Glu Lys Ala Glu Ala Ala Ala Thr
 25 30 35 40
 35 cta aag gca gcc cca ggc tgg cta aag cgg ttc ctg gta tgg aaa cct 196
 Leu Lys Ala Ala Pro Gly Trp Leu Lys Arg Phe Leu Val Trp Lys Pro
 45 50 55
 agg ccc gcg agt gcc cgg gcc cag ccc ggc cta gtt cag gaa gcg gct 244
 Arg Pro Ala Ser Ala Arg Ala Gln Pro Gly Leu Val Gln Glu Ala Ala
 60 65 70
 40 cag ccc cag ggc agc aca tca gag aca cca tgg aac aca gcc att cct 292
 Gln Pro Gln Gly Ser Thr Ser Glu Thr Pro Trp Asn Thr Ala Ile Pro
 75 80 85
 ctg ccg tcg tgc tgg gac cag tct ttc ctg acc aat atc acc ttc ttg 340
 45 Leu Pro Ser Cys Trp Asp Gln Ser Phe Leu Thr Asn Ile Thr Phe Leu
 90 95 100
 aag gtt ctt ctc tgg ttg gtc ctg ctg gga ctg ttt gtg gaa ctg gaa 388
 Lys Val Leu Leu Trp Leu Val Leu Leu Gly Leu Phe Val Glu Leu Glu
 105 110 115 120
 50 ttt ggc ctg gca tat ttt gtc ctg tcc ttg ttc tat tgg atg tac gtc 436
 Phe Gly Leu Ala Tyr Phe Val Leu Ser Leu Phe Tyr Trp Met Tyr Val
 125 130 135
 ggg aca cga ggc cct gaa gag aag aaa gag gga gag aag agc gcc tac 484
 Gly Thr Arg Gly Pro Glu Glu Lys Lys Glu Gly Glu Lys Ser Ala Tyr
 140 145 150
 55 tct gtg ttc aat cca ggc tgt gaa gcc atc cag ggc acc ctg act gca 532
 Ser Val Phe Asn Pro Gly Cys Glu Ala Ile Gln Gly Thr Leu Thr Ala
 155 160 165
 gag cag ttg gag cgc gag tta cag ttg aga ccc ctg gca ggg aga 577
 60 Glu Gln Leu Glu Arg Glu Leu Gln Leu Arg Pro Leu Ala Gly Arg
 170 175 180
 taggacccag ctgtgctgtc atgcagctaa cctctgatgt ggtcttcctc accattggct 637
 atggatttga tttcagggtgt ataggactaa gggcagcttg cgggttagct ctgtgactgc 697

```

atagtttttc taccttcttt cctgatctt ttgctgccat ttgatctttg atagtttttg 757
tgaaactctc taaaatacat tcaactgtggg tccgacgcaa tttataaaaa ttatgtactc 817
aagaagggag acctgtttgt ttcatctctc atctgttttg gagatgattt tagagcacta 877
gaaaggcact ggggagattc tcagcttaaa acatccagca gtttgaagta tgattaggta 937
5 catcagggct gcattgtcaa tgttctcttt aagtctttta acatttatag caattttttt 997
tttcccggag agtttaggtt gcaagttttg ggtttcttgt ttgtttttgt tttgcttcct 1057
gctttaattc ttttaattttc agtcattact ggtattgaaa aataaaatat ctttaaaaca 1117
tctcctcttc agaaataggt cctcttctcat tgcccatcac catcttccac tctcctatta 1177
ttttgccact actcagtaaa ggaaggtagg aagagacaaa cgcctaagtg caggtgtggg 1237
10 gagggatttc acaagtgggt attaacggcc agttcagcaa gaagtgttga gtgtgtacaa 1297
aggggagggc tggaagtgtt aactccagac ccgttggtg cttgagttgt ttcttatatt 1357
ctaaagcagc agtccctaac ctttttggca ccagggacca gttttgtgga acacagtttt 1417
tccatggacg ggggtgggtt ggaggatgaa acttccacct cagatcatca ggcattagag 1477
tctcataagg agcacgcaac ctagatccct cgcattgcga gttcacaata cggttctaag 1537
15 ggcttttaga taagcagctt tttcacctgt gggcctctgg tgagaaattc tgtaaatgtt 1597
gataatcagg ctggatttta atgctgcttt tccagtacaa tgtttagagt tgggttcatt 1657
aaaattaggc aaactcccat tgggttaggg cttctctcat tccattttgt ggctaacctt 1717
actgtgttcc agcccttgct gaaaattctt ctgatatgtg ttgcccttcc tcacagccct 1777
ttggccattg ggagtttggc tgtccctcag agccatccgg tcaagcagat ggtctgttct 1837
20 atctcacaga aaagtctttt cttccatgag ttctgtctga actgaacatg taaaaagtat 1897
gggaaacaga tgaatcccta ttaaacatga agttttgatt gtattcaaaa aaaaaaaaaa 1957
aaa 1960

```

<210> 180
 25 <211> 1443
 <212> DNA
 <213> Homo sapiens

<220>
 30 <221> CDS
 <222> 23..451

```

<400> 180
accggcgggc ggggcgggta ag atg gcg gcc ccg cgg cga ggg aga gga tcc 52
35 Met Ala Ala Pro Arg Arg Gly Arg Gly Ser
1 5 10
tcc aca gtg ctc tct tca gtt ccc ctt caa atg ctg ttt tat ctc agc 100
Ser Thr Val Leu Ser Ser Val Pro Leu Gln Met Leu Phe Tyr Leu Ser
15 20 25
40 gga acg tac tac gcc ctg tat ttc ctc gcc acg ctc ctg atg atc acg 148
Gly Thr Tyr Tyr Ala Leu Tyr Phe Leu Ala Thr Leu Leu Met Ile Thr
30 35 40
tat aaa agt cag gtg ttc agc tat cct cac cgc tac ctg gtc ctc gat 196
Tyr Lys Ser Gln Val Phe Ser Tyr Pro His Arg Tyr Leu Val Leu Asp
45 45 50 55
ctt gct ctg ctg ttt ctg atg ggg att cta gaa gca gtt cgg tta tac 244
Leu Ala Leu Leu Phe Leu Met Gly Ile Leu Glu Ala Val Arg Leu Tyr
60 65 70
ctg ggc acc agg ggc aac ctg aca gag gct gag agg ccg ctg gcc gcc 292
Leu Gly Thr Arg Gly Asn Leu Thr Glu Ala Glu Arg Pro Leu Ala Ala
75 80 85 90
agc ctg gcc ctc acg gct ggc acc gcc ctc ctc tct gcc cac ttc ctg 340
Ser Leu Ala Leu Thr Ala Gly Thr Ala Leu Ser Ala His Phe Leu
95 100 105
55 ctt tgg cag gcc cta gtg ttg tgg gcg gac tgg gcc ctc agc gcc acg 388
Leu Trp Gln Ala Leu Val Leu Trp Ala Asp Trp Ala Leu Ser Ala Thr
110 115 120
ctc ctg gcc ctt cac ggc ctg gag gcc gtc ctg cag gtg gtt gcc atc 436
Leu Leu Ala Leu His Gly Leu Glu Ala Val Leu Gln Val Val Ala Ile
125 130 135
60 gcg gcc ttc acc agg tagctacgga caccgggat accccacact ggggccctcc 491
Ala Ala Phe Thr Arg
140

```

```

tcctgggcct gaccagtcct ccagctgtca cctccccatt cctggacagg aagggcactt 551
ttcctagtga ctggccatag atgggttttg atggttccat ctgttctggc aggagtggga 611
gcaggagcca gggcagaaca aactgctgga ggccctgggtg ttgggaacag ctgcggggag 671
ggtagggacc agacagaact gccttcaaga tgagtccag gagcgcacac tcagccctgt 731
5 cagtgggggtc tggcttttagc agccaggcct ccacagaccc ccatgggccc ccagggccga 791
gagggaggac agagcccttc agaacagagg cctcatctca ctgcatcccc catcaccccc 851
tagttcccca atggtcctaa tttgtgttct gagatccag tttactctgt ggccaggccc 911
cacctgtgtt tccaagtcgg gctggagacg caggatgggg taggccttgt gctctgagca 971
acccagctc tgcctcacag gcaggcaggc ccgggtgcaag agtggactct gggttcctaa 1031
10 agcaataaat gcaaacaagc caacagctct gctgcctagc aatttccatc ttagccacac 1091
ttctcccttc aggggcttcg gaggagaggt cagggtcaag gccggggatg atactgcagg 1151
agagagagca gcggaggggc acattcggag cctccgtcca ctccagtttt atcagctttt 1211
gcttttgcac ggagtgtctaa acaaatctta gctctgtgtt tttttcccat tcccagattt 1271
actatcagtt ctcttataaa agtatctaag ctgttacagt agctttccct tcacttgatt 1331
15 ctattgtgtg ttttctatgt ttggaataat tacacccaaa tatctagata ttttctcttc 1391
accgcatttt gtaaataaag agatgtgtat gcctcaaaaa aaaaaaaaaa aa 1443

```

<210> 181

<211> 605

20 <212> DNA

<213> Homo sapiens

<220>

<221> CDS

25 <222> 232..450

<400> 181

```

caaatacaaa tgccccaaga agactgagga taggagaaag aatatctcta cctgtgaaac 60
attgtagtac tgcctggcta ggagttcatt gttgttttct gaaggacgta accaaccact 120
30 ccaaaactta caggcttaaa acaacaaaca tgtatcattt cttatgattc tgtgggttgg 180
ctgggtgggtt cttctagctg aggcaggatg gtctaggata gctacatcca c atg tct 237
Met Ser
1

```

```

ggg gtc cca gct gag atg act ggg gct gtt gag gcc ttt ctc cct gtg 285
35 Gly Val Pro Ala Glu Met Thr Gly Ala Val Glu Ala Phe Leu Pro Val
5 10 15

```

```

gtg tca tcc tcc aga agg ctg ccc aga ttt gtc cat atg gta gca gga 333
Val Ser Ser Ser Arg Arg Leu Pro Arg Phe Val His Met Val Ala Gly
20 25 30

```

```

gtt tcc tcg aag caa gag agg gca aga tcc aac aca gaa gca ctt ttc 381
40 Val Ser Ser Lys Gln Glu Arg Ala Arg Ser Asn Thr Glu Ala Leu Phe
35 40 45 50

```

```

aag ctc tgt ttc cat cac att tgc caa tgt ctc act gat gaa cac aag 429
Lys Leu Cys Phe His His Ile Cys Gln Cys Leu Thr Asp Glu His Lys
45 55 60 65

```

```

ttc cat ggc caa gtc cag ttt taagaaatgg agaaataggg cttgggtcag 480
Phe His Gly Gln Val Gln Phe
70

```

```

tggctcatgt ctgtaatccc agcacttttg gagggccaagg catgcggatc atttgaggtc 540
50 aggagttcca gaccagcctg gccaacatgg tgaaaaccca tctctaccaa aaaaaaaaaa 600
aaaaa 605

```

<210> 182

<211> 1724

55 <212> DNA

<213> Homo sapiens

<220>

<221> CDS

60 <222> 758..1183

<400> 182

```

aactaaagcc gggagtccgg tgaacgggca gaagcagggc catgcccaag ccacccccaa 60

```

	gateccccctg	aacctgcacc	tccatcacga	cccattcagg	agcctccagg	agcccagaca	120
	ccagcccccc	accatggtaa	gtccttcaag	ggtgggatct	ggaagaggaa	agaggaggga	180
	caccagccag	gtggaggtgt	cctaaaaatg	accatcagaa	attgggggtga	ggggaggggc	240
	atggtggact	tctgtggggg	tggggtgtct	ctcagtgcat	ctcaggtgcc	tccagcatcc	300
5	cttaccaggg	agcaagctcc	catctgtagg	tgggtggggat	gccaggggtgg	tatccctgga	360
	tccaaggata	gggcaggacc	tggaagacag	aaggtggccc	agggagaatc	acagagtctg	420
	cagggacaag	gacatagcct	cctttgcttg	caaatataagg	gagccctttc	ccagtccagc	480
	ccagtctctc	gtctccctgt	gtagccttgg	gctagtctct	tcccctctct	tggccccggg	540
	tccccacagat	gtcatatttg	gaaatccgtc	tagatgcgga	agttgctctt	caggggtctt	600
10	tcagttgcaa	cattctcaag	gtctgtgggt	tctgccacag	agtcctcggc	tgagatggga	660
	agctatgtct	aacaagcgat	ggggtggatt	gacgccctcc	ctgtgccggg	gacgggcggg	720
	atggctgcag	cagaggcagg	agaggctgaa	tacgtcc atg	cca ccc ttt	ggg ggg	775
				Met Pro Pro Phe Gly Gly			
				1	5		
15	cat ccc tta tcc caa gag gag gat	ggc agc cag agg tgt tgc tgc ctg	823				
	His Pro Leu Ser Gln Glu Glu Asp	Gly Ser Gln Arg Cys Cys Cys Leu					
		10	15	20			
	tca agt ctg agg tct gtc gat gat	agc aac ggg gag act gtc gtg atc	871				
	Ser Ser Leu Arg Ser Val Asp	Ser Asn Gly Glu Thr Val Val Ile					
20		25	30	35			
	atg gcg cta ttc cta gca gta tgc tac	cac cat aag acg caa agt aag	919				
	Met Ala Leu Phe Leu Ala Val Ser Tyr	His His Lys Thr Gln Ser Lys					
		40	45	50			
	agg tgg cca ggg ctg acc cca ccc	cac agc tct ctg ctg tgt aga cca	967				
25	Arg Trp Pro Gly Leu Thr Pro Pro	His Ser Ser Leu Leu Cys Arg Pro					
		55	60	65	70		
	ctt cag ctt tca ttt ctc gtc att	cag tca gtg agg atg aga gca tgt	1015				
	Leu Gln Leu Ser Phe Leu Val Ile	Gln Ser Val Arg Met Arg Ala Cys					
		75	80	85			
30	ggc tgt gac agc ggc cac tgc agg	att ctt ggc agg tac agc tta cta	1063				
	Gly Cys Asp Ser Gly His Cys Arg	Ile Leu Gly Arg Tyr Ser Leu Leu					
		90	95	100			
	ggg tgg agt cag gga cat agg gca	aga ggc aga ggt ggt gtt agt ctg	1111				
	Gly Trp Ser Gln Gly His Arg Ala	Arg Gly Arg Gly Gly Val Ser Leu					
35		105	110	115			
	aga gac aac acc ttc ttt cag gaa	gcc agt gag ggc cag gga cag tgg	1159				
	Arg Asp Asn Thr Phe Phe Gln Glu	Ala Ser Glu Gly Gln Gly Gln Trp					
		120	125	130			
	ctc atg cct gta atc cca gca ttt	taggaggctg agacaggtag atcaçttgag	1213				
40	Leu Met Pro Val Ile Pro Ala Phe						
		135	140				
	gtcaggtgtt	cgagaccagc	ctggccaacg	tgggtgaaacc	tcgtctctac	taaaaaatac	1273
	aaaaaattaa	ctgggcgtgg	tggcacacgc	ctgtaatccc	agctacatat	gaggctgagg	1333
	caagagaata	acttgaaccc	aggaggcgga	gggtgcagtg	agctgagatc	ctgccgctgc	1393
45	actccagcct	gggtgacaga	gcacactccg	tctcaaaaaa	ggaaagctga	tgagaaattg	1453
	ggcatcccgg	aattcacacc	caaaccatca	gctggagctc	tgagactgtt	ggggtgggaa	1513
	ttcttccaag	atgagaagca	agccagggag	gctcaggtcc	tgggatgggc	agggctttga	1573
	tcaaaagaac	acaggaagtg	atttgctact	tgaagaaaag	gcaacccttc	cccaaggaag	1633
	ccctctgaaa	atgcttagtc	aacagtcggc	ttggcagaca	aggtctggga	ggggccaccc	1693
50	gtatcgcaga	ggacaaaaaa	aaaaaaaaaa	a			1724
	<210>	183					
	<211>	1686					
	<212>	DNA					
55	<213>	Homo sapiens					
	<220>						
	<221>	CDS					
	<222>	486..932					
60	<400>	183					
	cggctcactg	cagcctcttc	ctcccagtct	caagtgattc	tcctgtctca	gcctcctgag	60
	tggctgggat	tacaggtgtg	caccactacc	acttggtctaa	tttttatact	tttagtagag	120

	atgggggtttc	accatgtttg	ccaggctggc	cttgaactcc	tgacctcagg	tgatccgccc	180
	gcctcggcct	cccaaagtgc	tgggggttaca	ggcatgagcc	accgcacccg	gcccccttcc	240
	ttcgtcttag	tcaatcctat	cccacctctt	cttccaccag	tccccctacc	tgatgggtccc	300
	aacactttcat	catccaccac	ctcctggagg	gggtaccccg	aggtgctccg	ctgggggactc	360
5	tgctcattct	gggggtgcag	ttgacggctg	gtcgtgatct	ttcccgtaat	ctgtccccctc	420
	ttacggaacc	tagtctccgt	tctgtccatg	gccttcttct	ggacactgct	aggatccaga	480
	agagt atg tta tca att	ctc aag cct agg aga	agt cag gag tgg aga aca				530
	Met Leu Ser Ile	Leu Lys Pro Arg Arg	Ser Gln Glu Trp Arg Thr				
	1	5	10	15			
10	gct ctg aga aga tac	tgt tgt cca act	gat ctc cag gca cca	cgg agt			578
	Ala Leu Arg Arg Tyr	Cys Cys Pro Thr	Asp Leu Gln Ala Pro	Arg Ser			
		20	25	30			
	ccg gtc cct cca atc	agg aag gtc gga atc	tct gat gtc atc	gtt cat			626
	Pro Val Pro Pro Ile	Arg Lys Val Gly Ile	Ser Asp Val Ile	Val His			
15		35	40	45			
	gcc aac ctg gca acc	agt ttg aaa aaa	aac aca tgt aac	tgc cag gct			674
	Ala Asn Leu Ala Thr	Ser Leu Lys Lys	Asn Thr Cys Asn	Cys Gln Ala			
		50	55	60			
	gat ctc ttg tcc tgg	aga tcc tgg gtg	aat ggt atc tcc	tgc cac tgt			722
20	Asp Leu Leu Ser Trp	Arg Ser Trp Val	Asn Gly Ile Ser	Cys His Cys			
		65	70	75			
	ccc aac ctc aga cca	ttg tcc aaa agc	atc ttc agg gac	tcc aca tcc			770
	Pro Asn Leu Arg Pro	Leu Ser Lys Ser	Ile Phe Arg Asp	Ser Thr Ser			
		80	85	90	95		
25	ctc tgt tcc ctg tcc	cag cag agg ctg	tgt cct ctc cac	tca aag cct			818
	Leu Cys Ser Leu Ser	Gln Gln Arg Leu	Cys Pro Leu His	Ser Lys Pro			
		100	105	110			
	gaa gca tgt tgg ggt	ctc ttt gtc tct	gta cat gcc cat	ttc aga gtc			866
	Glu Ala Cys Trp Gly	Leu Phe Val Ser	Val His Ala His	Phe Arg Val			
30		115	120	125			
	cag gct ggt ggg aga	ggg aac aga gtg	gga aag aaa act	agg gta agc			914
	Gln Ala Gly Gly Arg	Gly Asn Arg Val	Gly Lys Lys Thr	Arg Val Ser			
		130	135	140			
	aga aac gat gaa acc	tta taagagt	gag attatcatgt	gcaagagtga			962
35	Arg Asn Asp Glu Thr	Leu					
		145					
	gattatcatg	tacaagagat	cccaggaaat	actgactttg	atgaaaaagt	cacatcagag	1022
	cactcagttt	tggcagagct	ttttctgccc	aatgttttact	cacattcact	gtccgagatt	1082
	ctatactggg	ggtaacacag	tctctgccc	taaggcaatt	ttgagtccaa	gagacatttt	1142
40	gaggcctaaa	aatcatagga	aactgcccct	gagctcacac	atattttccaa	tggtgtcccc	1202
	aatttcaggg	aatccatgga	ttacctaaagc	cagccccctcc	agttcggcta	agaaactcta	1262
	gtctatatgt	caagttttgt	atcatatgta	ttgctctgaa	ctcagaaaatt	tcccttccat	1322
	ttatggattc	tatgaataaa	atatcacatg	tacaaaaaga	ctaagtcaaa	aaatttcagc	1382
	tgtgcacagt	ggctcatgct	tgtaatccca	gcactttggg	tggccgaggg	gggaggattg	1442
45	cctgaggcca	gcagttcaag	accagtatgg	gcaacatggc	aagagcccat	ctctaaaaaa	1502
	acaaaaccaa	accaaattgg	ccaggtgtgg	tggctggcac	ctgtgttcca	actacttggg	1562
	agactcatgt	gacaggaaga	tcacttgagc	ccgggggtta	gaggctgcag	tgagctatga	1622
	tcttgccact	gcactccagc	ctgggtgaca	gagcgagaca	ccgtcgcaaa	aaaaaaaaaa	1682
	aaaa						1686
50							
	<210>	184					
	<211>	463					
	<212>	DNA					
	<213>	Homo sapiens					
55							
	<220>						
	<221>	CDS					
	<222>	80..304					
60							
	<400>	184					
	cttttaacag	ctgaggtctc	tctttaattc	tcttaaatat	cattttctccc	tcaaaaaaga	60
	ccattagatc	atttcacaa	atg tat ctg	cca cca aac	agg tca gag	ctt tgc	112
			Met Tyr Leu	Pro Pro Asn	Arg Ser Glu	Leu Cys	

193

acagcaatga gcatgaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 771
aa 773

<210> 186
5 <211> 753
<212> DNA
<213> Homo sapiens

<220>
10 <221> CDS
<222> 94..573

<400> 186
acttttcagg ggacattcag aggcattcag cccttcctcc tcaccagctc ccagagttcc 60
15 catctccatc cccaatccta aagaaggaaa tcg atg cca cgg tcc tca agg agc 114
Met Pro Arg Ser Ser Arg Ser
1 5
cct ggg gac cca ggc gcc cta ctc gaa gat gtg gcc cac aat ccc aga 162
Pro Gly Asp Pro Gly Ala Leu Leu Glu Asp Val Ala His Asn Pro Arg
20 10 15 20
ccc cgg agg att gcc cag cga ggc cgg aac acc agc agg atg gca gag 210
Pro Arg Arg Ile Ala Gln Arg Gly Arg Asn Thr Ser Arg Met Ala Glu
25 25 30 35
gac acc tcc cca aac atg aat gac aac atc ctg ttg cct gtc cgc aac 258
Asp Thr Ser Pro Asn Met Asn Asp Asn Ile Leu Leu Pro Val Arg Asn
40 45 50 55
aat gac caa gcc cta ggc ctg act cag tgc atg ctg gga tgt gtg tcc 306
Asn Asp Gln Ala Leu Gly Leu Thr Gln Cys Met Leu Gly Cys Val Ser
60 65 70
30 tgg ttc acc tgt ttt gcc tgc tcc ctg aga act cag gcc cag cag gtt 354
Trp Phe Thr Cys Phe Ala Cys Ser Leu Arg Thr Gln Ala Gln Gln Val
75 80 85
ctg ttt aac acg tgc aga gac aga gtt tca cca tgt tgc cca ggc tgg 402
Leu Phe Asn Thr Cys Arg Asp Arg Val Ser Pro Cys Cys Pro Gly Trp
90 95 100
35 tct caa act cca gtg atc ctc cca cct cag cct tcc gaa gtg ctg gga 450
Ser Gln Thr Pro Val Ile Leu Pro Pro Gln Pro Ser Glu Val Leu Gly
105 110 115
tta cag atg caa gct gct gtg cca gaa gct cat gga gaa gac agg cat 498
Leu Gln Met Gln Ala Ala Val Pro Glu Ala His Gly Glu Asp Arg His
120 125 130 135
tct gct cct ctg tgc ttt cgg tgt gtc cca ggg ccc tgc cca gtc cca 546
Ser Ala Pro Leu Cys Phe Arg Cys Val Pro Gly Pro Cys Pro Val Pro
140 145 150
45 ggt gga ggt atc cct ggg ccc tgg cac tgattatagg aactgggca 593
Gly Gly Gly Ile Pro Gly Pro Trp His
155 160
agacactgca ctgccactg actcagtttc cccatctgcc tgatgggtgt tgctgtgaga 653
attatgaaat gaaatgatga ccatgaaaat attgtagaag ccaagaaatg cttcagaagt 713
50 tataaagctc tccccaaacc gtgttaaaaa aaaaaaaaaa 753

<210> 187
<211> 754
<212> DNA
55 <213> Homo sapiens

<220>
<221> CDS
<222> 181..462

60 <400> 187
atcctatcaa aagttacggt gaagtcaggg tgggtggcga gtccctgcaa ggtcgccct 60
ctgtgccaac acagcctgat ggcttcttgt ttcaggaaac atccagaatt acaactggcc 120

	attgagttat tacatatcaa ttgaacaagg tagtttttaa atgaaagaaa atcttgcaac	180
	atg aat aaa gag ata gac tct ttg aat ctg gca tac agc ttt ccc ttc	228
	Met Asn Lys Glu Ile Asp Ser Leu Asn Leu Ala Tyr Ser Phe Pro Phe	
	1 5 10 15	
5	ctt ctt cct gct ttc ctg gac aca ccg tgg aca gac cca ttt ccc tct	276
	Leu Leu Pro Ala Phe Leu Asp Thr Pro Trp Thr Asp Pro Phe Pro Ser	
	20 25 30	
	gga ttc atg gta agg tcc cga gtg ctt ctg ata cag ctg ctg agc aga	324
	Gly Phe Met Val Arg Ser Arg Val Leu Leu Ile Gln Leu Leu Ser Arg	
10	35 40 45	
	ccc cgc tca tct cag gag tcc cga gga cac tcg ctt ccc tgc agc ccg	372
	Pro Arg Ser Ser Gln Glu Ser Arg Gly His Ser Leu Pro Cys Ser Pro	
	50 55 60	
	tcc gcc ctc cat aag cct ggg ggc atc tgc cct gca gca ctg ggg agg	420
15	Ser Ala Leu His Lys Pro Gly Gly Ile Cys Pro Ala Ala Leu Gly Arg	
	65 70 75 80	
	agc cac ctc ctt gtc tgg gaa cag cca agc ctc cgt gac agc	462
	Ser His Leu Leu Val Trp Glu Gln Pro Ser Leu Arg Asp Ser	
	85 90	
20	tgaggattct tgtggattgt tctttctgta actggacagc acatccggaa ttccttgcca	522
	tagctctgtg ccttgctggg gtctgagggt cacaggctcag atgctgctgt ctggctcttc	582
	ccaattgcgg cgtgaattcc ttcactctca ccagtagctt cttgctctcc ccaaggagg	642
	cacgtgctta gtagggagag aggcctacca aggttgccat ctgccatggg ctcaattgtg	702
	tccccaaccc ccttgcaaat tatatatattga agtcccaaaa aaaaaaaaaa aa	754
25		
	<210> 188	
	<211> 998	
	<212> DNA	
	<213> Homo sapiens	
30		
	<220>	
	<221> CDS	
	<222> 6..290	
35		
	<220>	
	<221> misc_feature	
	<222> 871	
	<223> n=a, g, c or t	
40		
	<400> 188	
	gattc atg aag gcc tcg ggt cct gac ctc tct gat gga ctc cac tgc ccc	50
	Met Lys Ala Ser Gly Pro Asp Leu Ser Asp Gly Leu His Cys Pro	
	1 5 10 15	
	agt cta att aga cat tta aga acc ttc tct gca gct gct gcc tta gcc	98
45	Ser Leu Ile Arg His Leu Arg Thr Phe Ser Ala Ala Ala Leu Ala	
	20 25 30	
	cca aga tac cca acc aga ctt ccc agt tca ctg ctt cta tgg cac ctc	146
	Pro Arg Tyr Pro Thr Arg Leu Pro Ser Ser Leu Leu Leu Trp His Leu	
	35 40 45	
50	tgc cag tgc ctc cat ctc ctc tat gca gtt tct acc tca tgc aac agc	194
	Cys Gln Cys Leu His Leu Leu Tyr Ala Val Ser Thr Ser Cys Asn Ser	
	50 55 60	
	cat ggg aag aga tcg gct gcc tgg gca atg acc aga aca gaa gac aca	242
	His Gly Lys Arg Ser Ala Ala Trp Ala Met Thr Arg Thr Glu Asp Thr	
55	65 70 75	
	gat gcg cta aca gat tcc ttc gat gac agt ttc atc agt tct gca gat	290
	Asp Ala Leu Thr Asp Ser Phe Asp Asp Ser Phe Ile Ser Ser Ala Asp	
	80 85 90 95	
	taaagacttt caccagaaaa aaaaattacc tgattttgcc ctgaggcagc cagggagggc	350
60	tttgtccttg acaatcccac tgacttatatt aacaggtagc tcaaaaccca acaaaaactg	410
	gaggaggctg ctccactgca gggatgggtt caattcggtg actggagtat tgtactctcc	470
	ttgcaccctg gctcatcccc acaaaagacc tttcaaagaa aacacttaat tacctccttg	530
	cacaagccct gtaagcccta aggtgaaaaa aaactcagca gacaagggtcc acagagaagg	590

agaaggcaca attcagtagg gacctacgct cagcaccagg ataaagaaac tgtccattcc 650
 tgccacctcc taggaagcta aaagaattaa ggggaggccg ggcacggtgg ctcacgcctg 710
 taatcccagc actttgggag gccgaggcgg gtggatcatg aggtcaggag atcgagacca 770
 tcctggctaa catggtgaaa ccccatctct actaaaaata caaaaaatta gccgggctg 830
 5 gtggcgggcg ccctgtagtc ccagctactc gggaggctga nggcaggaga atggtgtgaa 890
 cctgggaggc ggagcttgca gtgagccgag attgcgcct gctccactcc agcctgagcg 950
 acagagcgag actccgtctc aaaaaaaaaa argaaaaaaaa aaaaaaaaaa 998

<210> 189
 10 <211> 605
 <212> DNA
 <213> Homo sapiens

<220>
 15 <221> CDS
 <222> 115..411

<400> 189
 aagaaagggg tgaggcctaa gggacaatca ggatgttttt cagagagaag tgtggatgct 60
 20 ggacaggaag aaccacagat accagatacg ggtactgttg taactctgtt ctcc atg 117
 Met
 1
 aaa aaa aag gaa gaa aca aca ctt tca gag atg gag cct gtt gag cca 165
 Lys Lys Lys Glu Glu Thr Thr Leu Ser Glu Met Glu Pro Val Glu Pro
 25 5 10 15
 cag tac caa cta gtc aat gct gaa tcg act tct ccc ttt cta cat tgc 213
 Gln Tyr Gln Leu Val Asn Ala Glu Ser Thr Ser Pro Phe Leu His Cys
 20 25 30
 ctg aga gaa gtc att ggg gaa tac tct gta cac gaa ttt tca ctg ttg 261
 30 Leu Arg Glu Val Ile Gly Glu Tyr Ser Val His Glu Phe Ser Leu Leu
 35 40 45
 ggg aaa aca gag agt caa ggg att gga ttg tgg att gca ttg gtg gtt 309
 Gly Lys Thr Glu Ser Gln Gly Ile Gly Leu Trp Ile Ala Leu Val Val
 50 55 60 65
 35 ttc ctc agt ttc ctc atc ttc tcc aca agt ttc tac ata tcg aat gca 357
 Phe Leu Ser Phe Leu Ile Phe Ser Thr Ser Phe Tyr Ile Ser Asn Ala
 70 75 80
 gag cag ccc ttc ttc aaa gaa cct cct acg gaa gct gct aag gaa ctc 405
 Glu Gln Pro Phe Phe Lys Glu Pro Pro Thr Glu Ala Ala Lys Glu Leu
 40 85 90 95
 agt ctg tagctctgcg tggagccatg tgtaaacact gaactgagac ctgccacctc 461
 Ser Leu
 ctactaccta agggccatt ttcactctgat atcatccccc agaaacaaac tcatgatgac 521
 ttccatgttt ttttttagatt agatacatgg agaattttcc tttcccttag aattaaaatc 581
 45 ctgcattcta aaaaaaaaaa aaaa 605

<210> 190
 <211> 526
 <212> DNA
 50 <213> Homo sapiens

<220>
 <221> CDS
 <222> 3..368

55 <400> 190
 ag atc cga gcg acc atg gtg gcc cgg gtg tgg tcg ctg atg agg ttc 47
 Ile Arg Ala Thr Met Val Ala Arg Val Trp Ser Leu Met Arg Phe
 1 5 10 15
 60 ctc atc aag gga agt gtg gct ggg ggc gcc gtc tac ctg gtg tac gac 95
 Leu Ile Lys Gly Ser Val Ala Gly Gly Ala Val Tyr Leu Val Tyr Asp
 20 25 30
 cag gag ctg ctg ggg ccc agc gac aag agc cag gca gcc cta cag aag 143

	Gln	Glu	Leu	Leu	Gly	Pro	Ser	Asp	Lys	Ser	Gln	Ala	Ala	Leu	Gln	Lys	
				35					40					45			
	gct	ggg	gag	gtg	gtc	ccc	ccc	gcc	atg	tac	cag	ttc	agc	cag	tac	gtg	191
	Ala	Gly	Glu	Val	Val	Pro	Pro	Ala	Met	Tyr	Gln	Phe	Ser	Gln	Tyr	Val	
5			50					55				60					
	tgt	cag	cag	aca	ggc	ctg	cag	ata	ccc	cag	ctc	cca	gcc	cct	cca	aag	239
	Cys	Gln	Gln	Thr	Gly	Leu	Gln	Ile	Pro	Gln	Leu	Pro	Ala	Pro	Pro	Lys	
			65				70					75					
	att	tac	ttt	ccc	atc	cgt	gac	tcc	tgg	aat	gca	ggc	atc	atg	acg	gtg	287
10	Ile	Tyr	Phe	Pro	Ile	Arg	Asp	Ser	Trp	Asn	Ala	Gly	Ile	Met	Thr	Val	
						85					90					95	
	atg	tca	gct	ctg	tcg	gtg	gcc	ccc	tcc	aag	gcc	cgc	gag	tac	tcc	aag	335
	Met	Ser	Ala	Leu	Ser	Val	Ala	Pro	Ser	Lys	Ala	Arg	Glu	Tyr	Ser	Lys	
					100				105						110		
15	gag	ggc	tgg	gag	tat	gtg	aag	gcg	cgc	acc	aag	tagcgagtc	gcaggggccc				388
	Glu	Gly	Trp	Glu	Tyr	Val	Lys	Ala	Arg	Thr	Lys						
				115				120									
	cctgccccgg	ccagaacggg	cagggctgcc	actgacctga	agactccgga	ctgggacccc											448
	actccgaggg	cagctccccg	ccttgccggc	ccaataaagg	acttcagaag	tgaaaaaaaa											508
20	ataaaaaaaaa	aaaaaaaaa															526
	<210>	191															
	<211>	910															
	<212>	DNA															
25	<213>	Homo sapiens															
	<220>																
	<221>	CDS															
	<222>	174..527															
30	<400>	191															
	attttcctgt	taggccaaga	gagaagagga	tccttctctca	gagcctccag	cctcccttga											60
	tcccttgctt	gtgggcatat	gtgggtcata	tttccctccc	atcacctct	gcacgccacc											120
	cccatcaccg	ccacagaccc	ccagcccttc	agttgccctg	cacctccttg	gtg atg											176
35																Met	
																1	
	cag	ccg	tcc	ttg	tta	agg	tca	tac	agg	ttg	aag	gcc	caa	tta	agc	ctg	224
	Gln	Pro	Ser	Leu	Leu	Arg	Ser	Tyr	Arg	Leu	Lys	Ala	Gln	Leu	Ser	Leu	
				5					10					15			
40	tca	tct	aca	gtt	ccc	cga	aga	atc	acg	gac	aaa	cca	gcc	aca	aag	tcc	272
	Ser	Ser	Thr	Val	Pro	Arg	Arg	Ile	Thr	Asp	Lys	Pro	Ala	Thr	Lys	Ser	
				20				25				30					
	tgg	gaa	gga	ggc	agg	agg	gag	ctg	tgt	cct	cgg	gta	ctc	ttc	acc	caa	320
	Trp	Glu	Gly	Gly	Arg	Arg	Glu	Leu	Cys	Pro	Arg	Val	Leu	Phe	Thr	Gln	
45				35			40				45						
	ctc	ctt	ctc	tgg	gtt	tgg	cct	gga	gat	cct	ggc	cct	gaa	ctc	cag	gaa	368
	Leu	Leu	Leu	Trp	Val	Trp	Pro	Gly	Asp	Pro	Gly	Pro	Glu	Leu	Gln	Glu	
				50		55				60					65		
	aca	ggc	ttc	cct	ggc	cca	cct	cgc	cca	gct	cac	ctc	aaa	act	gac	cga	416
50	Thr	Gly	Phe	Pro	Gly	Pro	Pro	Arg	Pro	Ala	His	Leu	Lys	Thr	Asp	Arg	
				70				75						80			
	gcc	atc	atg	gtt	ggc	gtc	aaa	ggc	att	gaa	gag	aaa	agt	ggc	ata	ggc	464
	Ala	Ile	Met	Val	Gly	Val	Lys	Gly	Ile	Glu	Glu	Lys	Ser	Gly	Ile	Gly	
				85				90						95			
55	gct	gga	gtc	tgc	agg	gtg	agt	gtg	gag	aag	ttg	gct	tcc	aca	cag	gag	512
	Ala	Gly	Val	Cys	Arg	Val	Ser	Val	Glu	Lys	Leu	Ala	Ser	Thr	Gln	Glu	
				100				105				110					
	agg	act	tcc	tcc	ctc	taaggagctc	cccatacccc	ccatcacctt	ggcattccca								567
	Arg	Thr	Ser	Ser	Leu												
60				115													
	gctcctccag	aatccctccc	tcctcagcc	tagagaagga	caactgcttc	cccttggggc											627
	ttgtcccctc	acctccttga	ggaaagaact	gggagtaaat	ctgcttgaag	ttctcctcat											687
	tgacaattcc	gctgggacat	tcctggaagg	agagggcacc	aggctgaggg	cagagacaaa											747

```

atcccccttcc gttcaccgcc cccaccctcc atggcccaag actcccaggg aggggggataa 807
tcttcaagcc tccagaggac tcaccacgtg gctcatgtga tgggagggaa gacttcttcc 867
ccagtgcaca aataaaaaaac atggaacgaa aaaaaaaaaa aaa 910

5 <210> 192
  <211> 668
  <212> DNA
  <213> Homo sapiens

10 <220>
   <221> CDS
   <222> 57..203

   <400> 192
15 tcctgtcgac gtgttcttcc ggtggcggag cggcggatta gccttcgcgg ggcaaa atg 59
                                   Met
                                   1
gag ctc gag gcc atg agc aga tat acc agc cca gtg aac cca cct gtc 107
Glu Leu Glu Ala Met Ser Arg Tyr Thr Ser Pro Val Asn Pro Pro Val
20      5      10      15
ttc ccc cat ctg acc gtg gtg ctt ttg gcc att ggc atg ttc ttc acc 155
Phe Pro His Leu Thr Val Val Leu Leu Ala Ile Gly Met Phe Phe Thr
      20      25      30
gcc tgg ttc ttc gtg tat cct ttc act gag cag cca gag gac cag cat 203
Ala Trp Phe Phe Val Tyr Pro Phe Thr Glu Gln Pro Glu Asp Gln His
      35      40      45
tagtgatgtg ggaagctcag ggagaaacca cgctaggtac atggaccccg cgggttttgt 263
acattggatt ggggctgaga gaagattgcc gtgggctggg ctctctgcac tccacagtcc 323
acccttcgcg tttgccttaa ctgctgtgcc cagttacgag gtcacctcta ccaagtacac 383
30 tcgtgatata tataaagagc tcctcatctc attagtggcc tcaactctta tgggctttgg 443
agtcctcttc ctgctgctct gggttggcat ctacgtgtga gcacccaagg gtaacaacca 503
gatggcttca ctgaaacctg cttttgtaaa ttactttttt ttactgttgc tggaagtgtc 563
ccacctgctg ctcataataa atgcagatgt atagcaaaaa aaaaaaaaaa aaaaaaaaaa 623
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaa 668

35 <210> 193
   <211> 637
   <212> DNA
   <213> Homo sapiens

40 <220>
   <221> CDS
   <222> 68..334

45 <400> 193
agttatgaag ttctaaaagc aagtcttaat caggaagtgt ccttgatcac caacggctcg 60
cccaggc atg ctg gct ctc ttc cac ttc cac ctt cca cca tgg gat gac 109
      Met Leu Ala Leu Phe His Phe His Leu Pro Pro Trp Asp Asp
      1      5      10
gca gta aga agg cca tca gta gat gcc agt ccc tca acc ttg aac ttt 157
Ala Val Arg Arg Pro Ser Val Asp Ala Ser Pro Ser Thr Leu Asn Phe
      15      20      25      30
cca gac gca gaa ctt tat gcc tcc att ttc ctc tgc tgc atg gcc cca 205
Pro Asp Ala Glu Leu Tyr Ala Ser Ile Phe Leu Cys Cys Met Ala Pro
      35      40      45
gga gag att tta att agc ttt cta acc ttg gtc cag att gca cat gca 253
Gly Glu Ile Leu Ile Ser Phe Leu Thr Leu Val Gln Ile Ala His Ala
      50      55      60
aat ggt aga gga tgc aac acc ccc gct tgt gga gct gcc gct tgt gtc 301
60 Asn Gly Arg Gly Cys Asn Thr Pro Ala Cys Gly Ala Ala Ala Cys Val
      65      70      75
tgg cat gaa aat tca caa gaa gag agg aaa tac tgaggagaaa atggcagatt 354
Trp His Glu Asn Ser Gln Glu Glu Arg Lys Tyr

```

80 85

gtggtttgctg aatttgattg acgaagaagt caccatgaaa atcacagtga accattttgga 414
aagcaaaactg ccaaaaaaat aatagttagt catgctctca ggctgggttg tttggctgtt 474
gtgggtttct tgcattttcca gatgattgca aagagctgtt tctcaatttc tgcaacaagt 534
5 gccagctgaa attttggtac cagtttcatt aaatatgtat aacaaaaakaa aaaaaaaaaa 594
aaaaaaaaaa aaaaaaaaaa aaaaaaagaa aaaaaaaaaa aaa 637

<210> 194
<211> 706
10 <212> DNA
<213> Homo sapiens

<220>
<221> CDS
15 <222> 183..443

<400> 194

agaagttctc agaggggtgag ggtcccatcat ctcttcgagg acaggcccta gctaccgagt 60
cacagaaacc cagggccgaa gcaaagtccc aatcccagag aggctggggc acacctacaa 120
20 ctgaaaggag gcttagaaat ccttcagaga ccaccctatc gggtctcctc cacctggaca 180
gg atg agc cag caa cac aga agg aag agg cct tcc tcc gaa aga aaa 227
Met Ser Gln Gln His Arg Arg Lys Arg Pro Ser Ser Glu Arg Lys
1 5 10 15

agc aca aga aag atg gac aca tgg cag agt ctt aaa gtc aaa gaa gta 275
25 Ser Thr Arg Lys Met Asp Thr Trp Gln Ser Leu Lys Val Lys Glu Val
20 25 30

ttc tgt aag cat aat tct tcc tat gaa tgc ctt ctc tat aaa gag gtt 323
Phe Cys Lys His Asn Ser Ser Tyr Glu Cys Leu Leu Tyr Lys Glu Val
35 40 45

30 gaa gca aga cag gtt tct aag aca gcc acc gat ggg tcc tac ctc ctc 371
Glu Ala Arg Gln Val Ser Lys Thr Ala Thr Asp Gly Ser Tyr Leu Leu
50 55 60

gta ttc aca tcc tat gta atc tcc tcc cca gtg tgg act gga cct ggt 419
Val Phe Thr Ser Tyr Val Ile Ser Ser Pro Val Trp Thr Gly Pro Gly
35 65 70 75

gac ttg ctt cca gtg aat aga ata tagcaaaagt gattgatgtc acctccaaga 473
Asp Leu Leu Pro Val Asn Arg Ile

80 85

ttcagctata gaagactatg actatgactt tcctcttggc tagcattctc gctaaccctt 533
40 cctgcttgct tgtactgagc tgccctatga agaggcccat gtaggggtggc ctgggtgggg 593
gtgatctgtg gccaacagcc agcaaggaac taaatcctgt ttacaaccac atgagcttgg 653
aaggagatcc ttccccagta aagccaggag atgaatacaa aaaaaaaaaa aaa 706

<210> 195
45 <211> 670
<212> DNA
<213> Homo sapiens

<220>
50 <221> CDS
<222> 94..228

<400> 195

acttttcagg ggacattcag aggcatcagc cccttcctcc tcaccagctc ccagagttcc 60
55 catctccatc cccaatccta aagaaggaaa tcg atg cca cgg tcc tca agg agc 114
Met Pro Arg Ser Ser Arg Ser
1 5

cct ggg gac cca ggc gcc cta ctc gaa gat ggc cca caa tcc cag acc 162
Pro Gly Asp Pro Gly Ala Leu Leu Glu Asp Gly Pro Gln Ser Gln Thr
10 15 20

60 ccg gag gat tgc cca gcg agg ccg gaa cac cag cag gat ggc aga gga 210
Pro Glu Asp Cys Pro Ala Arg Pro Glu His Gln Asp Gly Arg Gly
25 30 35

cac ctc ccc aaa cat gaa tgacaacatc ctgttgccctg tccgcaacaa 258
 His Leu Pro Lys His Glu
 40 45
 5 tgaccaagcc ctaggcctga ctcagtgcat gctgggatgt gtgtcctggt tcacctgttt 318
 tgccctgctcc ctgagaactc aggcccagca gggtctgttt aacacgtgca gatgcaagct 378
 gctgtgccag aagctcatgg agaagacagg cattctgctc ctctgtgctt tccgtgtgtc 438
 ccagggccct gccagtcctc aggtggaagg tatccctggg ccctggcact gattatagga 498
 cactgggcaa gacactgcac cgccacgtga ctcagtttcc ccctctgcct gatgggtgtt 558
 gctgtgagaa ttatgaaatg aaatgatgac catgaaaata ttgtagaagc caagaaatgc 618
 10 ttcagaagtt ataaagctct ccccaaaccg tgttatgaaa aaaaaaaaaa aa 670

 <210> 196
 <211> 510
 <212> DNA
 15 <213> Homo sapiens

 <220>
 <221> CDS
 <222> 133..327
 20
 <400> 196
 aacctcaagg agccctgttg tgctaccgac tgcagagctc atggacatcc atcaggaagc 60
 ctccaatacc caaaccaggg gtagttgcct aatccatata catgtggata gctctttact 120
 taggaaacct tg atg gct tat ttg gat gac aaa ggt tcc ctt ttg gcg ata 171
 25 Met Ala Tyr Leu Asp Asp Lys Gly Ser Leu Leu Ala Ile
 1 5 10
 cat agc cat gcg aga caa cat agc cat gaa aca aac caa gtc cac cag 219
 His Ser His Ala Arg Gln His Ser His Glu Thr Asn Gln Val His Gln
 15 20 25
 30 tgg ctt cct agg aac aca ttt gct ttc ctg ata aaa gag gac aga tgc 267
 Trp Leu Pro Arg Asn Thr Phe Ala Phe Leu Ile Lys Glu Asp Arg Cys
 30 35 40 45
 agt tgc aga agt acc tgt gcc tct ttt tct tct tct tct ttt tct 315
 Ser Cys Arg Ser Thr Cys Ala Ser Phe Ser Phe Ser Ser Ser Phe Ser
 35 50 55 60
 ttt tta atc tct taaatgcaga tataagaact ggtactgaag cagccatctt 367
 Phe Leu Ile Ser
 65
 40 gtgaccataa ggaagaagcc aagaacatca gaaccagtgg cctagccatt gcacagtcac 427
 ctaaacacac ctctggactt gttattatgt aaaaaaaaaa aaacacctgc tcttgttatt 487
 tgcaatccaa aaaaaaaaaa aaa 510

 <210> 197
 <211> 500
 45 <212> DNA
 <213> Homo sapiens

 <220>
 <221> CDS
 50 <222> 22..357

 <400> 197
 atagaatata cacaaaagga a atg aga aag aaa tgt aaa tgc ttc act ata 51
 Met Arg Lys Lys Cys Lys Cys Phe Thr Ile
 5 10
 55 aaa aaa aca aat aca tac gaa gaa agt aat gca gga aat gaa gga caa 99
 Lys Lys Thr Asn Thr Tyr Glu Glu Ser Asn Ala Gly Asn Glu Gly Gln
 15 20 25
 aaa gaa gct ata agc att tgt att tgc aga aga gat ggt tta ctt cct 147
 60 Lys Glu Ala Ile Ser Ile Cys Ile Cys Arg Arg Asp Gly Leu Leu Pro
 30 35 40
 ctg tgg gta acc agg tta tca gat ttg gtg ttt tcc aaa gaa aag gca 195
 Leu Trp Val Thr Arg Leu Ser Asp Leu Val Phe Ser Lys Glu Lys Ala

```

      45              50              55
cat ggc atg att cca ctt ctt ggc tcc cat agg gaa aag aag aca agt 243
His Gly Met Ile Pro Leu Leu Gly Ser His Arg Glu Lys Lys Thr Ser
      60              65              70
5  aaa gag atg aag act tct tcc agg aac ctg agg tac ttc att gtc tgc 291
Lys Glu Met Lys Thr Ser Ser Arg Asn Leu Arg Tyr Phe Ile Val Cys
      75              80              85              90
aga gat gcc tca tcc tac acc cct cag tca ctc ata tct gga tac att 339
Arg Asp Ala Ser Ser Tyr Thr Pro Gln Ser Leu Ile Ser Gly Tyr Ile
10  gga cct tgt caa cat caa taatggacat acctctgata tttgaactct 387
Gly Pro Cys Gln His Gln
      110
gaatctcact ctgtgaccac aactttgtat ctttctaagt ctttaatctt caacctcaca 447
15  gaactcttca taccctaaaa tatagtattt tcacctggaa aaaaaaaaaa aaa 500

<210> 198
<211> 667
<212> DNA
20  <213> Homo sapiens

<220>
<221> CDS
<222> 4..333
25  <400> 198
aaa atg gtg ttt gga gcc atg gtc ctt ctt gtg gga ctt gaa gaa ctg 48
Met Val Phe Gly Ala Met Val Leu Leu Val Gly Leu Glu Glu Leu
      1      5      10      15
30  acc aat atc cgc aac gtg gag aga ctg aag aag gac ttg agg gcc agt 96
Thr Asn Ile Arg Asn Val Glu Arg Leu Lys Lys Asp Leu Arg Ala Ser
      20      25      30
tat tgc ctc atc gac agc ttc ctg ggg gac tcg gag ctc atc ggg gac 144
Tyr Cys Leu Ile Asp Ser Phe Leu Gly Asp Ser Glu Leu Ile Gly Asp
35  ctg acc cag tgt gtg gac tgc gtg att cct cca gag ggg tcc ctc ttg 192
Leu Thr Gln Cys Val Asp Cys Val Ile Pro Pro Glu Gly Ser Leu Leu
      40      45      50      55      60
40  cag atc tct agc tac ctc tac tta aat act gct ctt gtg gac ttg cct 240
Gln Ile Ser Ser Tyr Leu Tyr Leu Asn Thr Ala Leu Val Asp Leu Pro
      65      70      75
ggg gtg gcg gcc tcc cag gca tgt gac tct cag cag gtg act tgg ctt 288
Gly Val Ala Ala Ser Gln Ala Cys Asp Ser Gln Gln Val Thr Trp Leu
      80      85      90      95
45  ctc tac gtt gct aat ggt gcc tac tcg gca tgt aac agg cct gga 333
Leu Tyr Val Ala Asn Gly Ala Tyr Ser Ala Cys Asn Arg Pro Gly
      100      105      110
tgaacggtag ctgctgcggt tacattatta gcttcagttt gccccgccag gctagatggt 393
taatcagatt tcacagactt cacagtgtga gttggggatg tgacttcgta tgaaagtga 453
50  ggaactcagg ctgagagagg gtgagacgta ggagcatggc cactgcgcga gctcggggct 513
ggctgtgggt ttctcccat tccctgccca tctgggaagt cgctgccacc ccctacgctt 573
gtctgctgac tcccagtcct cctaaccctc cagaatgtaa acagcagcag atgaacaaaa 633
ataaaaaatac aaaaggccga aaaaaaaaaa aaaa 667

55  <210> 199
<211> 514
<212> DNA
<213> Homo sapiens

60  <220>
<221> CDS
<222> 1..363

```

<400> 199		
	acg agt tct tcc ggg gcg gag gtc acc atg gca gct gcc ttg gct cgg	48
	Thr Ser Ser Ser Gly Ala Glu Val Thr Met Ala Ala Leu Ala Arg	
	1 5 10 15	
5	ctt ggt ctg cgg cct gtc aaa cag gtt cgg gtt cag ttc tgt ccc ttc	96
	Leu Gly Leu Arg Pro Val Lys Gln Val Arg Val Gln Phe Cys Pro Phe	
	20 25 30	
	gag aaa aac gtg gaa tcg acg agg acc ttc ctg cag acg gtg agc agt	144
	Glu Lys Asn Val Glu Ser Thr Arg Thr Phe Leu Gln Thr Val Ser Ser	
10	35 40 45	
	gag aag gtc cgc tcc act aat ctc aac tgc tca gtg att gcg gac gtg	192
	Glu Lys Val Arg Ser Thr Asn Leu Asn Cys Ser Val Ile Ala Asp Val	
	50 55 60	
	agg cat gac ggc tcc gag ccc tgc gtg gac gtg ctg ttc gga gac ggg	240
15	Arg His Asp Gly Ser Glu Pro Cys Val Asp Val Leu Phe Gly Asp Gly	
	65 70 75 80	
	cat cgc ctg att atg cgc ggc gct cat ctc acc gct ctg gaa atg ctc	288
	His Arg Leu Ile Met Arg Gly Ala His Leu Thr Ala Leu Glu Met Leu	
	85 90 95	
20	acc gcc ttc gcc tcc cac atc cgg gcc agg gac gcg gcg ggc agc ggg	336
	Thr Ala Phe Ala Ser His Ile Arg Ala Arg Asp Ala Ala Gly Ser Gly	
	100 105 110	
	gac aag ccg ggc gct gat act ggt cgc tgacagcgcc aaagagacca	383
	Asp Lys Pro Gly Ala Asp Thr Gly Arg	
25	115 120	
	acaagatgat ttgcgtggac taggacactt aacctaagaa gagtttctact taatcattca	443
	aatcactatc tgaagggtca cggagcgcaa aataaagttt aaaaccctgc taccaaaaaa	503
	aaaaaaaaa a	514
30 <210> 200		
<211> 462		
<212> DNA		
<213> Homo sapiens		
35 <220>		
<221> CDS		
<222> 41..337		
<400> 200		
40	cttcaccacc aaaactctcc actccaccag cacagccaaa atg ctc gca cgt gct	55
	Met Leu Ala Arg Ala	
	1 5	
	act ttc cgc gcc gcc tcg gcc cca act ctc gtc gcc cgc cgc ggc ttc	103
	Thr Phe Arg Ala Ala Ser Ala Pro Thr Leu Val Ala Arg Arg Gly Phe	
45	10 15 20	
	cag tcg acc cgc gcg caa atg gcc agc cca tac cac tac ccc gag ggt	151
	Gln Ser Thr Arg Ala Gln Met Ala Ser Pro Tyr His Tyr Pro Glu Gly	
	25 30 35	
	cct cgc agc aac ttg cca ttc gac ccg ctg aag aag ggc ttt gct ttc	199
50	Pro Arg Ser Asn Leu Pro Phe Asp Pro Leu Lys Lys Gly Phe Ala Phe	
	40 45 50	
	aag tac tgg ggc ttt atg ggc acc gga ttc gcc ctt ccc ttc ctc ctt	247
	Lys Tyr Trp Gly Phe Met Gly Thr Gly Phe Ala Leu Pro Phe Leu Leu	
	55 60 65	
55	gct gtc tgg caa aca gaa caa gcc gta aat gcg ctg aga cac ggc gtg	295
	Ala Val Trp Gln Thr Glu Gln Ala Val Asn Ala Leu Arg His Gly Val	
	70 75 80 85	
	gac atg cgt atc ggg atc ccg ggg aac acg gca ttt gta gat	337
	Asp Met Arg Ile Gly Ile Pro Gly Asn Thr Ala Phe Val Asp	
60	90 95	
	taggtggagg gcccgcatac ggctatacta gacatcacag catcaatttc attgtctgtc	397
	ccccaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa	457
	aaaaa	462

<210> 201
 <211> 551
 <212> DNA
 5 <213> Homo sapiens

<220>
 <221> CDS
 <222> 1..549

10 <400> 201
 aga gag gga gcc cga gcc agg cca tct cca acc atg tcc gac gag gcc 48
 Arg Glu Gly Ala Arg Ala Arg Pro Ser Pro Thr Met Ser Asp Glu Ala
 1 5 10 15
 15 tcg gcc atc act tcc tac gag aag ttt cta acc ccc gag gag ccc ttc 96
 Ser Ala Ile Thr Ser Tyr Glu Lys Phe Leu Thr Pro Glu Glu Pro Phe
 20 25 30
 cca ctc ctg gga cct cct cgc ggg gtg ggc acc tgc ccg agc gag gag 144
 Pro Leu Leu Gly Pro Pro Arg Gly Val Gly Thr Cys Pro Ser Glu Glu
 20 35 40 45
 ccg ggc tgc ctg gac atc agc gac ttc ggc tgc cag ctg tcc tcc tgc 192
 Pro Gly Cys Leu Asp Ile Ser Asp Phe Gly Cys Gln Leu Ser Ser Cys
 50 55 60
 cat cgc acc gac ccg ctc cac cgc ttc cac acc aac agg tgg aac cta 240
 25 His Arg Thr Asp Pro Leu His Arg Phe His Thr Asn Arg Trp Asn Leu
 65 70 75 80
 act tct tgt gga aca agt gtt gcc agc tca gaa ggc agt gag gag ctg 288
 Thr Ser Cys Gly Thr Ser Val Ala Ser Ser Glu Gly Ser Glu Glu Leu
 85 90 95
 30 ttt tca tct gtg tct gtt gga gat caa gat gat tgc tat tcc ctg tta 336
 Phe Ser Ser Val Ser Val Gly Asp Gln Asp Asp Cys Tyr Ser Leu Leu
 100 105 110
 gat gat cag gac ttc act tct ttt gat tta ttt cct gag ggg agt gtc 384
 Asp Asp Gln Asp Phe Thr Ser Phe Asp Leu Phe Pro Glu Gly Ser Val
 115 120 125
 35 tgc agt gat gtc tct tct tct att agc act tac tgg gat tgg tca gat 432
 Cys Ser Asp Val Ser Ser Ser Ile Ser Thr Tyr Trp Asp Trp Ser Asp
 130 135 140
 agc gag ttt gaa tgg cag tta cca ggc agt gac att gcc agt ggg agt 480
 40 Ser Glu Phe Glu Trp Gln Leu Pro Gly Ser Asp Ile Ala Ser Gly Ser
 145 150 155 160
 gat gta ctt tct gat gtc ata ccc agt att cca agt tca cct tgc ctg 528
 Asp Val Leu Ser Asp Val Ile Pro Ser Ile Pro Ser Ser Pro Cys Leu
 165 170 175
 45 ctt cct aaa aaa aaa aaa aa 551
 Leu Pro Lys Lys Lys Lys Lys
 180

<210> 202
 50 <211> 550
 <212> DNA
 <213> Homo sapiens

<220>
 55 <221> CDS
 <222> 34..315

<220>
 <221> misc_feature
 60 <222> 483
 <223> n=a, g, c or t

<400> 202

	agagaggggag cccgagccag gccatctcca acc atg tcc gac gag gcc tcg gcc	54
	Met Ser Asp Glu Ala Ser Ala	
	1 5	
5	atc act tcc tac gag aag ttt cta acc ccc gag gag ccc ttc cca ctc	102
	Ile Thr Ser Tyr Glu Lys Phe Leu Thr Pro Glu Glu Pro Phe Pro Leu	
	10 15 20	
	ctg gga cct cct cgc ggg gtg ggc acc tgc ccg agc gag gag ccg ggc	150
	Leu Gly Pro Pro Arg Gly Val Gly Thr Cys Pro Ser Glu Glu Pro Gly	
	25 30 35	
10	tgc ctg gac atc agc gac ttc ggc tgc cag ctg tcc tcc tgc cat cgc	198
	Cys Leu Asp Ile Ser Asp Phe Gly Cys Gln Leu Ser Ser Cys His Arg	
	40 45 50 55	
	acc gac ccg ctc cac cgc ttc cac acc aac agg tgg aac cta act tct	246
	Thr Asp Pro Leu His Arg Phe His Thr Asn Arg Trp Asn Leu Thr Ser	
15	60 65 70	
	tgt gga aca agt gtt gcc agc tca gaa ggc agt gag gag ctg ttt tca	294
	Cys Gly Thr Ser Val Ala Ser Ser Glu Gly Ser Glu Glu Leu Phe Ser	
	75 80 85	
20	tct gtc tgt tgg aga tca aga tgattgctat tccctgtag atgatcagga	345
	Ser Val Cys Trp Arg Ser Arg	
	90	
	cttcacttct tttgatttat ttcctgaggg gagtgtctgc agtgatgtct cttcttctat	405
	tagcacttac tgggattggg cagatagcga gtttgaatgg cagttaccag gcagtgcacat	465
	tgccagtggg agtgatgnta ctttctgatg tcataccag tattccaagt tcaccttgcc	525
25	tgcttcctaa aaaaaaaaaa aaaaa	550
	<210> 203	
	<211> 408	
	<212> DNA	
30	<213> Homo sapiens	
	<220>	
	<221> CDS	
	<222> 1..315	
35	<400> 203	
	atc ggg ccg cga gcg ccc tcc ccg tcg ttt tcc gtg aga gac gta gag	48
	Ile Gly Pro Arg Ala Pro Ser Pro Ser Phe Ser Val Arg Asp Val Glu	
	1 5 10 15	
40	ctg agc gac cca gcc cgc gag cga ggt gag atg ccg gtg gcc gtg ggt	96
	Leu Ser Asp Pro Ala Arg Glu Arg Gly Glu Met Pro Val Ala Val Gly	
	20 25 30	
	ccc tac gga cag tcc cag cca agc tgc ttc gac cgt gtc aaa atg ggc	144
	Pro Tyr Gly Gln Ser Gln Pro Ser Cys Phe Asp Arg Val Lys Met Gly	
45	35 40 45	
	ttc gtg atg ggt tgc gcc gtg ggc atg gcg gcc ggg gcg ctc ttc ggc	192
	Phe Val Met Gly Cys Ala Val Gly Met Ala Ala Gly Ala Leu Phe Gly	
	50 55 60	
	acc ttt tcc tgt ctc agg atc gga atg cgg ggt cga gag ctg atg ggc	240
50	Thr Phe Ser Cys Leu Arg Ile Gly Met Arg Gly Arg Glu Leu Met Gly	
	65 70 75 80	
	ggc att ggg aaa acc atg atg cag agt ggc ggc acc ttt ggc aca ttc	288
	Gly Ile Gly Lys Thr Met Met Gln Ser Gly Gly Thr Phe Gly Thr Phe	
	85 90 95	
55	atg gcc att ggg atg ggc atc cga tgc taaccatggg tgccaactac	335
	Met Ala Ile Gly Met Gly Ile Arg Cys	
	100 105	
	atctgtccct tcccatcaat cccagcccat gtactaataa aagaaagtct ttgagcaaaa	395
	aaaaaaaaaa aaa	408
60	<210> 204	
	<211> 665	
	<212> DNA	

<213> Homo sapiens

<220>

<221> CDS

5 <222> 94..582

<400> 204

```

acttttcagg ggacattcag aggcattcag cccttcctcc tcaccagctc ccagagttcc 60
catctccatc cccaatccta aagaaggaaa tcg atg cca cgg tcc tca agg agc 114
10 Met Pro Arg Ser Ser Arg Ser
      1 5
cct ggg gac cca ggc gcc cta ctc gaa gat gtg gcc cac aat ccc aga 162
Pro Gly Asp Pro Gly Ala Leu Leu Glu Asp Val Ala His Asn Pro Arg
      10 15 20
15 ccc cgg agg att gcc cag cga ggc cgg aac acc agc agg atg gca gag 210
Pro Arg Arg Ile Ala Gln Arg Gly Arg Asn Thr Ser Arg Met Ala Glu
      25 30 35
gac acc tcc cca aac atg aat gac aac atc ctg ttg cct gtc cgc aac 258
Asp Thr Ser Pro Asn Met Asn Asp Asn Ile Leu Leu Pro Val Arg Asn
20 40 45 50 55
aat gac caa gcc cta ggc ctg act cag tgc atg ctg gga tgt gtg tcc 306
Asn Asp Gln Ala Leu Gly Leu Thr Gln Cys Met Leu Gly Cys Val Ser
      60 65 70
25 tgg ttc acc tgt ttt gcc tgc tcc ctg aga act cag gcc cag cag gtt 354
Trp Phe Thr Cys Phe Ala Cys Ser Leu Arg Thr Gln Ala Gln Gln Val
      75 80 85
ctg ttt aac acg tgc aga tgc aag ctg ctg tgc cag aag ctc atg gag 402
Leu Phe Asn Thr Cys Arg Cys Lys Leu Leu Cys Gln Lys Leu Met Glu
      90 95 100
30 aag aca ggc att ctg ctc ctc tgt gct ttc ggt gtg tcc cag ggc cct 450
Lys Thr Gly Ile Leu Leu Leu Cys Ala Phe Gly Val Ser Gln Gly Pro
      105 110 115
gcc cag tcc cag gtg gag gta tcc ctg ggc cct ggc act gat tat agg 498
Ala Gln Ser Gln Val Glu Val Ser Leu Gly Pro Gly Thr Asp Tyr Arg
35 120 125 130 135
aca ctg ggc aag aca ctg cac tgc cac gtg act cag ttt ccc cat ctg 546
Thr Leu Gly Lys Thr Leu His Cys His Val Thr Gln Phe Pro His Leu
      140 145 150
40 cct gat ggg tgt tgc tgt gag aat tat gaa atg aaa tgatgacat 592
Pro Asp Gly Cys Cys Cys Glu Asn Tyr Glu Met Lys
      155 160
gaaaatattg tagaagccaa gaaatgcttc agaagttata aagctctccc caaaccgcaa 652
aaaaaaaaaa aaa 665

```

45 <210> 205

<211> 1008

<212> DNA

<213> Homo sapiens

50 <220>

<221> CDS

<222> 540..923

<400> 205

```

55 atttaggtga gctgccacgt ccggaggagg gcagcaagaa tgaaagacct ctagttttcc 60
agactcccgg agccctgggtc tctacaccac atggacgtta tccacctcct ctgtgtcctc 120
ccaaggcagc atttcagaag gtgatccacg gcaaagccgt cccttcaaat ccgtctttgt 180
gcccactgcc atagtcaacc ccgtgagaag cacagccggc cctgggactt taggacaagg 240
gtctcttcgg aaagggcgga gcagcatgag aaagagtaag tgggtggcaga gagatggatc 300
60 cctgcagaga cccctccagt ccgggatccc cactctcgtg gtaggctccc tcagacgcag 360
ccccaccatg gtccttcggc ctcagcagtt ccaattctac cagccacagg ggatcacctc 420
ctccccctca gccgtgggtg tggagatggg gtccaagcct gccctcacgg gggagcccgc 480
cctcacgtgc atcagcaggg gcagtgaggc ggatccactc cgcggccagc tccctcatt 539

```

atg gaa gac aaa gaa atc ccc atc aag agt gag cct ctg cca aaa ccg 587
 Met Glu Asp Lys Glu Ile Pro Ile Lys Ser Glu Pro Leu Pro Lys Pro
 1 5 10 15
 ccc gca tct gcc cca cca tcc atc ctg gtg aaa cca gaa aac tca aga 635
 5 Pro Ala Ser Ala Pro Pro Ser Ile Leu Val Lys Pro Glu Asn Ser Arg
 20 25 30
 aat gga atc gaa aag caa gtc aaa acc gtg aga ttt cag aat tac agc 683
 Asn Gly Ile Glu Lys Gln Val Lys Thr Val Arg Phe Gln Asn Tyr Ser
 35 40 45
 cct cct ccc acc aaa cat tac acc tcc cat ccc acc tcc gga aag cct 731
 10 Pro Pro Pro Thr Lys His Tyr Thr Ser His Pro Thr Ser Gly Lys Pro
 50 55 60
 gaa cag cca gcc acc ctc aag gcg tcc cag cct gaa gca gcg tcc ttg 779
 Glu Gln Pro Ala Thr Leu Lys Ala Ser Gln Pro Glu Ala Ala Ser Leu
 15 65 70 75 80
 ggc cca gag atg acc gtc cta ttt gcc cac cga agt ggc tgc cac tcc 827
 Gly Pro Glu Met Thr Val Leu Phe Ala His Arg Ser Gly Cys His Ser
 85 90 95
 gga cag cag aca gac ctc cgg aga aag tca gct ctt gcc aag gcc aca 875
 20 Gly Gln Gln Thr Asp Leu Arg Arg Lys Ser Ala Leu Ala Lys Ala Thr
 100 105 110
 acc ctg gtg tcc act gcc tca ggc acg cag acc gtg ttt ccc agc aaa 923
 Thr Leu Val Ser Thr Ala Ser Gly Thr Gln Thr Val Phe Pro Ser Lys
 115 120 125
 25 tgaacctacg ggtggctttt cctagacccc aaagaggtga attgcattta aatacagtct 983
 gcctycactg aaaaaaaaaa aaaaa 1008

 <210> 206
 <211> 455
 30 <212> DNA
 <213> Homo sapiens

 <220>
 <221> CDS
 35 <222> 77..364

 <400> 206
 tggacaaatg gacctgcggt aggagagagg gacaacagta ggagcaggca gatcttgctg 60
 tttcaaccaa aacctc atg ctg acc aga gtt gag gaa cag aag aag atg gtg 112
 40 Met Leu Thr Arg Val Glu Glu Gln Lys Lys Met Val
 1 5 10
 aag gcc tgc agg tat agg tgt tca gca tgt cat ctg aaa tat tcc cca 160
 Lys Ala Cys Arg Tyr Arg Cys Ser Ala Cys His Leu Lys Tyr Ser Pro
 15 20 25
 45 cag agg caa aaa gaa agg aaa tta tct ctg aaa agg ggg agg aca agt 208
 Gln Arg Gln Lys Glu Arg Lys Leu Ser Leu Lys Arg Gly Arg Thr Ser
 30 35 40
 cag cag aat atg tca atg ttt tgg ttg aag aag ctg ctt gaa tct ggg 256
 Gln Gln Asn Met Ser Met Phe Trp Leu Lys Lys Leu Leu Glu Ser Gly
 50 45 50 55 60
 ctt ttc tgt gcc atg tgt tct ccc agg gcc agc aca aag aag ggc ttt 304
 Leu Phe Cys Ala Met Cys Ser Pro Arg Ala Ser Thr Lys Lys Gly Phe
 65 70 75
 tgg tgc agg ccc aag acc acc ata atc atc att gat tat tcc tct cca 352
 55 Trp Cys Arg Pro Lys Thr Thr Ile Ile Ile Ile Asp Tyr Ser Ser Pro
 80 85 90
 cgc cag tgt ctc taaataaact ttctcttctt tctctgaaaa aaaaaaaaaa 404
 Arg Gln Cys Leu
 95
 60 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaagaaaa aaaaaaaaaa a 455

 <210> 207
 <211> 749

<212> DNA

<213> Homo sapiens

<220>

5 <221> CDS

<222> 65..544

<400> 207

```

ctttttacgac ggcgcggaaa gcaacggcaa gggccgcagc cagcaccggg cggagagggc      60
10 tacc atg ggg aaa atc gcg ctg caa ctc aaa gcc acg ctg gag aac atc      109
    Met Gly Lys Ile Ala Leu Gln Leu Lys Ala Thr Leu Glu Asn Ile
        1         5         10        15
    acc aac ctc cgg ccc gtg ggc gag gac ttc cgg tgg tac ctg aag atg      157
    Thr Asn Leu Arg Pro Val Gly Glu Asp Phe Arg Trp Tyr Leu Lys Met
        20        25        30
    15 aaa tgt ggc aac tgt ggt gag att tcg gac aag tgg cag tac atc cgg      205
    Lys Cys Gly Asn Cys Gly Glu Ile Ser Asp Lys Trp Gln Tyr Ile Arg
        35        40        45
    ctg atg gac agt gtg gca ctg aag ggg ggc cgt ggc agt gct tcc atg      253
    20 Leu Met Asp Ser Val Ala Leu Lys Gly Gly Arg Gly Ser Ala Ser Met
        50        55        60
    gtc cag aag tgc aag ctg tgt gca aga gaa aat tcc atc gag att tta      301
    Val Gln Lys Cys Lys Leu Cys Ala Arg Glu Asn Ser Ile Glu Ile Leu
        65        70        75
    25 agc agc acc atc aag cct tac aat gct gaa gac aat gag aac ttc aag      349
    Ser Ser Thr Ile Lys Pro Tyr Asn Ala Glu Asp Asn Glu Asn Phe Lys
        80        85        90        95
    aca ata gtg gag ttt gag tgc cgg ggc ctt gaa cca gtt gat ttc cag      397
    Thr Ile Val Glu Phe Glu Cys Arg Gly Leu Glu Pro Val Asp Phe Gln
        100       105       110
    30 ccg cas gwg rtw ttg ctg ctg aag gtg tgg agt cag gga cag cct tca      445
    Pro Xaa Xaa Xaa Leu Leu Leu Lys Val Trp Ser Gln Gly Gln Pro Ser
        115       120       125
    gtg aca tta atc tgc agg aga agg act ggg act gac tat gat gaa aag      493
    35 Val Thr Leu Ile Cys Arg Arg Arg Thr Gly Thr Asp Tyr Asp Glu Lys
        130       135       140
    gcc cag gag tct gtg gga atc tat gag gtc acc cac cag ttt gtg aag      541
    Ala Gln Glu Ser Val Gly Ile Tyr Glu Val Thr His Gln Phe Val Lys
        145       150       155
    40 tgc tgatccctct tccttcccag ttgcccttaa gaactgagaa aggacaaagt      594
    Cys
    160
    actctaagca gcagagccca cagaggctcg ttcctttgac ccttgtctcc tgggtggctat      654
    acgaaacctt cacaatctgc atgctggact ttattacagc ttcccaagcc ccatcaataa      714
    45 agccctgtt cagctacaa aaaaaaaaaa aaaaa      749

```

<210> 208

<211> 594

<212> DNA

50 <213> Homo sapiens

<220>

<221> CDS

<222> 117..467

55

<400> 208

```

aaatgtagcc tgggtggtgtt cccaggagga aaagaacgag agactggtgg cagcacaccc      60
tgggcccccc actccccgcc gcaagtcttg aggatggcca gcagagaaac aagaaa atg      119
                                                Met
                                                1
60 gac tcc ctg gct gct gga gag ttg aat gcc agc cac cag cca tgg gtg      167
    Asp Ser Leu Ala Ala Gly Glu Leu Asn Ala Ser His Gln Pro Trp Val
        5         10        15

```


	cca gag ttt gta gcc tat tgg agg aaa aca cac caa gat cac ctc tgc	215
	Pro Glu Phe Val Ala Tyr Trp Arg Lys Thr His Gln Asp His Leu Cys	
	20 25 30	
	agc ctg cac agc cgg gcc ttt gga ctc ctg gat gct aga gtg acc tgg	263
5	Ser Leu His Ser Arg Ala Phe Gly Leu Leu Asp Ala Arg Val Thr Trp	
	35 40 45	
	gcg ctg agg agg gcc ccc gag cca gta cca gga aag gat aga ctc ctg	311
	Ala Leu Arg Arg Ala Pro Glu Pro Val Pro Gly Lys Asp Arg Leu Leu	
	50 55 60 65	
10	ctt gca gca ttc cca gca gag gca tcg cct gtg gac acc gcg tct gtg	359
	Leu Ala Ala Phe Pro Ala Glu Ala Ser Pro Val Asp Thr Ala Ser Val	
	70 75 80	
	tct gta tat ggc aga gct ccc aga tat atg cac aag gga gtg aaa aaa	407
	Ser Val Tyr Gly Arg Ala Pro Arg Tyr Met His Lys Gly Val Lys Lys	
15	85 90 95	
	tgt gtt tgc acc cca gtc tct aaa aat tca aca gcc tgg tta ctt ctg	455
	Cys Val Cys Thr Pro Val Ser Lys Asn Ser Thr Ala Trp Leu Leu Leu	
	100 105 110	
	ggg ggt ata tcg taggtggcct taatacgtgt tatttgctca tctgtatttc	507
20	Gly Gly Ile Ser	
	115	
	ttactctttg cacaattaaa ccatgttccct tttacttatg tacattttta ataaaagaaa	567
	gttggttaacg aaaaaaaaaa aaaaaaa	594
25	<210> 209	
	<211> 2098	
	<212> DNA	
	<213> Homo sapiens	
30	<220>	
	<221> CDS	
	<222> 893..1897	
	<400> 209	
35	accaggtcct ccgtgggtgca cctgaaatgg tcaacagaag tcttgtgaca cgtggaatca	60
	tttcagagtc accccttctg cctcctgctc aagcaacaga cctgccgatc acccccgctg	120
	ggcccgcgtt tctcagggtc ttcctaatacc cctgggcttt ccggcttgct gtgtgcctgg	180
	agtcaggccg ccgtgcggca ggctgttaac cttagcctcg ggagagtggg atggagccac	240
	cttctcatgg aacgatcctc gccttccttc atctccattg ttttatggct tcacacggac	300
40	cgtggctttc tgcattggaag cttgggtggc aggggtgctgt cactttggga agcagccaga	360
	gaaccacagg atgcgtgaat cggctcctct gtcttcatgg gcatctccgg ccagggtggg	420
	gtcttgtgct gtgattagtg ggtcactggc aagtgtctga atgaagtgga ggttcgggtg	480
	gcaacagtga cggggaaggg ctatgggtcc gcctcaatgt catctgcccc atccctgggc	540
	ctccaggaat ccagggtctc agccctgct ttagaaggaa gtcctgacgg ccacgctgga	600
45	tctggaggac gtccggagct acagggcgga gatttcatct cgaaacctgg cggccagcag	660
	ggcgagcccc taccacagag tgaagggtgga ctttgccctc tcgtgccacg aggacttgct	720
	ggcaccatc tctgagccca tcgagtggaa ataccacagc cctgaggagg agataagcct	780
	tggacctgcc tgctggctct gggatttttt aagacgaagt caacaggcag ggtttttgct	840
	gcccttgagt ggcgggggtgg acagcgcagc caccgcctgc ctcactact cc atg tgc	898
50		
	Met Cys	
	1	
	tgc cag gtc tgc gag gcc gtg agg agt gga aat gag gaa gtg ctg gct	946
	Cys Gln Val Cys Glu Ala Val Ser Gly Asn Glu Glu Val Leu Ala	
	5 10 15	
55	gat gtc cgc acc atc gtg aac cag atc agc tac acc ccc cag gat ccc	994
	Asp Val Arg Thr Ile Val Asn Gln Ile Ser Tyr Thr Pro Gln Asp Pro	
	20 25 30	
	cga gac ctc tgt gga cgc ata ctg acc acc tgc tac atg gcc agc aag	1042
	Arg Asp Leu Cys Gly Arg Ile Leu Thr Thr Cys Tyr Met Ala Ser Lys	
60	35 40 45 50	
	aac tcc tcc cag gag acg tgc acc cgg gcc aga gag ttg gcc cag cag	1090
	Asn Ser Ser Gln Glu Thr Cys Thr Arg Ala Arg Glu Leu Ala Gln Gln	
	55 60 65	

	att gga agc cac cac atc agt ctc aac atc gat cca gcc gtg aag gcc	1138
	Ile Gly Ser His His Ile Ser Leu Asn Ile Asp Pro Ala Val Lys Ala	
	70 75 80	
5	gtc atg ggc atc ttc agc ctg gtg acg ggg aag agc cct ctg ttt gca	1186
	Val Met Gly Ile Phe Ser Leu Val Thr Gly Lys Ser Pro Leu Phe Ala	
	85 90 95	
	gct cat gga gga agc agc agg gaa aac ctg gcg ctg caa aat gtg cag	1234
	Ala His Gly Gly Ser Ser Arg Glu Asn Leu Ala Leu Gln Asn Val Gln	
	100 105 110	
10	gct cga ata cgg atg gtc ctc gcc tat ctg ttt gct cag ttg agc ctc	1282
	Ala Arg Ile Arg Met Val Leu Ala Tyr Leu Phe Ala Gln Leu Ser Leu	
	115 120 125 130	
	tgg tct cgg ggt gtc cac ggt ggg ctc ctc gtg ctg gga tcc gcc aac	1330
	Trp Ser Arg Gly Val His Gly Gly Leu Leu Val Leu Gly Ser Ala Asn	
15	135 140 145	
	gtg gat gag agt ctc ctg ggc tac ctg acc aag tac gac tgc tcc agt	1378
	Val Asp Glu Ser Leu Leu Gly Tyr Leu Thr Lys Tyr Asp Cys Ser Ser	
	150 155 160	
20	gcg gac atc aac ccc ata ggc ggg atc agc aag acg gac ctc agg gcc	1426
	Ala Asp Ile Asn Pro Ile Gly Gly Ile Ser Lys Thr Asp Leu Arg Ala	
	165 170 175	
	ttc gtc cag ttc tgc atc cag cgc ttc cag ctt cct gcc ctg cag agc	1474
	Phe Val Gln Phe Cys Ile Gln Arg Phe Gln Leu Pro Ala Leu Gln Ser	
	180 185 190	
25	atc ctg ttg gcg ccg gcc acc gca gag ctg gag ccc ttg gct gat gga	1522
	Ile Leu Leu Ala Pro Ala Thr Ala Glu Leu Glu Pro Leu Ala Asp Gly	
	195 200 205 210	
	cag gtg tcc cag acc gac gag gaa gat atg ggg atg aca tat gcg gag	1570
	Gln Val Ser Gln Thr Asp Glu Glu Asp Met Gly Met Thr Tyr Ala Glu	
30	215 220 225	
	ctc tcg gtc tat ggg aaa ctc agg aag gtg gcc aag atg ggg ccc tac	1618
	Leu Ser Val Tyr Gly Lys Leu Arg Lys Val Ala Lys Met Gly Pro Tyr	
	230 235 240	
35	agc atg ttc tgc aaa ctc ctc ggc atg tgg aga cac atc tgc acc ccg	1666
	Ser Met Phe Cys Lys Leu Leu Gly Met Trp Arg His Ile Cys Thr Pro	
	245 250 255	
	aga cag gtc gct gac aaa gtg aag cgg ttt ttc tcc aag tac tcc atg	1714
	Arg Gln Val Ala Asp Lys Val Lys Arg Phe Phe Ser Lys Tyr Ser Met	
	260 265 270	
40	aac aga cac aag atg acc acg ctc aca ccc gcg tac cac gcc gag aac	1762
	Asn Arg His Lys Met Thr Thr Leu Thr Pro Ala Tyr His Ala Glu Asn	
	275 280 285 290	
	tac agc cct gag gac aac agg ttt gat ctg cga cca ttt ctg tac aac	1810
	Tyr Ser Pro Glu Asp Asn Arg Phe Asp Leu Arg Pro Phe Leu Tyr Asn	
45	295 300 305	
	aca agc tgg cct tgg cag ttt cgg tgc ata gaa aat cag gtg cta cag	1858
	Thr Ser Trp Pro Trp Gln Phe Arg Cys Ile Glu Asn Gln Val Leu Gln	
	310 315 320	
50	ctc gag agg gca gag cca cag tcc ctg gac ggc gtg gac tgaggccggt	1907
	Leu Glu Arg Ala Glu Pro Gln Ser Leu Asp Gly Val Asp	
	325 330 335	
	tccttcctgg aggcctcctg tcctcgggga cccagcacc tcacatcag cattgctgga	1967
	gccaagggtta ggagccctac actaggagcc caggatggga cggcgcacga gccgagaggg	2027
	agggaaacttt tcagtcacaaat tcctcaaaaaa gaggctggaa taaagcctgg gctcaaaaaa	2087
55	aaaaaaaaa a	2098

<210> 210

<211> 428

<212> DNA

60 <213> Homo sapiens

<220>

<221> CDS

<222> 85..342

<400> 210

```

acactggtac agtcacctag cccatcagtt ccttcgtcga cagcgccggg gacatccaga      60
5 ctacaattta cagttcctct atcc atg tgc tgg gtt ata aat cat gcc atc      111
      Met Cys Trp Val Ile Asn His Ala Ile
      1               5
ctc cct aga atg aga atg cac agc aag cgg cag aca atc acc cgg cat      159
10 Leu Pro Arg Met Arg Met His Ser Lys Arg Gln Thr Ile Thr Arg His
      10      15      20      25
tcg gca tct ctt tct ttt cac gcg ctc cct cgc tcc gcc ttt ctc cag      207
Ser Ala Ser Leu Ser Phe His Ala Leu Pro Arg Ser Ala Phe Leu Gln
      30      35      40
ctc tgc ctt ctc agg cag ata cat cag ata cct tgt tta tcc atc ttc      255
15 Leu Cys Leu Leu Arg Gln Ile His Gln Ile Pro Cys Leu Ser Ile Phe
      45      50      55
agc tcc act ctg agg gcg cag acg cac gat tcc ggg atc ggg tgc acc      303
Ser Ser Thr Leu Arg Ala Gln Thr His Asp Ser Gly Ile Gly Cys Thr
      60      65      70
20 acg gcg aas cca ggc ggg aga cgg cag gag cag ctc agg taaccagggg      352
Thr Ala Xaa Pro Gly Gly Arg Arg Gln Glu Gln Leu Arg
      75      80      85
aagcttgctg gccacggag atgcagccgt ggagctgtga ggaaagacgg tctggcttca      412
aaaaaaaaa aaaaaa      428
25
<210> 211
<211> 769
<212> DNA
<213> Homo sapiens
30
<220>
<221> CDS
<222> 155..433
35 <400> 211
atttttcccc ccttgctcgg gatggtgccca caggaggctg tgcgggcccc gctccgcttc      60
gaatggtgga tgctgtgggg caccacctcc ttgaggacca aggcactcca gctgccagga      120
atttggtgc taacctcaca cagctgagcc ttcc atg aaa att gct ctc tgc caa      175
      Met Lys Ile Ala Leu Cys Gln
      1               5
40 aga gaa ctt cct agt cca agg tca tgt cta ctc tcc aga gat gtg act      223
Arg Glu Leu Pro Ser Pro Arg Ser Cys Leu Leu Ser Arg Asp Val Thr
      10      15      20
gga gtg att tgc acc cgg atg cct aga ctc gcc atc tgc tca aag act      271
45 Gly Val Ile Cys Thr Arg Met Pro Arg Leu Ala Ile Cys Ser Lys Thr
      25      30      35
gct cag aaa gcc ctc cca tgc att ccc ctg ctg cat acc agc cca ctc      319
Ala Gln Lys Ala Leu Pro Cys Ile Pro Leu Leu His Thr Ser Pro Leu
      40      45      50      55
50 tgc ctg cag ctg ctg tct gca gga ctt cat atc tat gcc aca ctg tgt      367
Cys Leu Gln Leu Leu Ser Ala Gly Leu His Ile Tyr Ala Thr Leu Cys
      60      65      70
aaa agc tgt gct tca aga aat cac aaa aac att ttc ctg cac cta cta      415
Lys Ser Cys Ala Ser Arg Asn His Lys Asn Ile Phe Leu His Leu Leu
      75      80      85
55 cac agc ctg agt gcg gca taagttgacc ttgcttgcta agaaatgggg      463
His Ser Leu Ser Ala Ala
      90
caagaaatgc ttttttgtat gtgtcatgtc tgtttgtttt tcaattaaga gaggaaagca      523
60 ttaggcagat ggaatgtaca tgtgaggatg aggagacaga aaacaagtag ccctttccat      583
caagatagag ggttttcttg ggttgctggc tattgaatgt cactcctgat ttctctttcc      643
aaggcactgt accaccagcc tactgagatt gtgtgggagc tttcatgggg gttgtatttc      703
actgatgaaa ataaattttt tgcataatgt gaaaaaaaaa aaaaaaaaaa aaaaaaaaaa      763

```

aaaaaa

769

<210> 212

<211> 914

5 <212> DNA

<213> Homo sapiens

<220>

<221> CDS

10 <222> 63..386

<400> 212

```

cttttttaggg agtccaaggt acagtcgccg cgtgcggagc ttgttactgg ttacttggcc 60
tc atg gcg gtc cga gct tcg ttc gag aac aac tgt gag atc ggc tgc 107
15 Met Ala Val Arg Ala Ser Phe Glu Asn Asn Cys Glu Ile Gly Cys
    1           5           10           15
    ttt gcc aag ctc acc aac acc tac tgt ctg gta gcg atc gga ggc tca 155
    Phe Ala Lys Leu Thr Asn Thr Tyr Cys Leu Val Ala Ile Gly Gly Ser
           20           25           30
20 gag aac ttc tac agt gtg ttc gag ggc gag ctc tcc gat acc atc ccc 203
    Glu Asn Phe Tyr Ser Val Phe Glu Gly Glu Leu Ser Asp Thr Ile Pro
           35           40           45
    gtg gtg cac gcg tct atc gcc ggc tgc cgc atc atc ggg cgc atg tgt 251
    Val Val His Ala Ser Ile Ala Gly Cys Arg Ile Ile Gly Arg Met Cys
25           50           55           60
    gtg gga gac aga aga aat tct ggc aga tgt gct caa ggt gga agt ctt 299
    Val Gly Asp Arg Arg Asn Ser Gly Arg Cys Ala Gln Gly Gly Ser Leu
           65           70           75
    cag aca gac agt ggc cga cca ggt gct agt agg aag cta ctg tgt ctt 347
30 Gln Thr Asp Ser Gly Arg Pro Gly Ala Ser Arg Lys Leu Leu Cys Leu
    80           85           90           95
    cag caa tca ggg agg gct ggt gca tcc caa gac ttc aat tgaagaccag 396
    Gln Gln Ser Gly Arg Ala Gly Ala Ser Gln Asp Phe Asn
           100           105
35 gatgagctgt cctctcttct tcaagtcgcc cttgtggcgg ggactgtgaa ccgaggcagt 456
    gaggtgattg ctgctgggat ggtgggtgaat gactggtgtg ccttctgtgg cctggacaca 516
    accagcacag agctgtcagt ggtggagagt gtcttcaagc tgaatgaagc ccagcctagc 576
    accattgccca ccagcatgcg ggattccctc attgacagcc tcacctgagt caccttccaa 636
    gttgttccat gggctcctgg ctctggactg tggccaacct tctccacatt ccgccaatc 696
40 tgtacctgat gctggcaggg aggtggcaga gagctcactg ggactgaggg gctgggcacc 756
    caaccctttt ccacctgtgc ttatcgccctg gatctatcat tactgcaaaa acctgctctg 816
    ttgtgctggc tggcaggccc tgtggctgct ggctgagggt tctgctgtcc tgtgccaccc 876
    cattaagtg cagttccctc caaaaaaaaa aaaaaaaaaa 914

```

45 <210> 213

<211> 1489

<212> DNA

<213> Homo sapiens

50 <220>

<221> CDS

<222> 460..1290

<400> 213

```

55 cttcttttccc tctccgtttt ggtgggctgg ttgaagatga aatccactga ggaggggaagt 60
    ccagcaccct gtgtgccagt ccagaactgg cccatctgta gacccctga aaatcatatg 120
    ggcttggatt tggatattct caacagaaag ggtaaaggc tgatggtacc taaagcctgg 180
    tacttgaatt ttgatcaaga taagctgcct taagttctct tcattacaca aatgatccta 240
    gataattgat agatcctgtg gttcaactgg atttctagat agaagctgga ttcattgtgat 300
60 gccagaggag taaaatttca agagactgaa accagatctg agtttcgctg ttccagtctg 360
    gacctctttg gtgctgtaaa tcttgatgat actgtagatg agtactgcgt ttttctttta 420
    tggactctct tcagcttctg gagacctcac tatectatt atg tct ttg tgt gaa 474
                                Met Ser Leu Cys Glu

```

```

    <210> 214
    <211> 776
60  <212> DNA
    <213> Homo sapiens

    <220>

```

<221> CDS

<222> 21..539

<400> 214

```

5  caaatatttc catcacgggg atg ctt gtc atg tac ctg ctt gcc gcc ctc ttt      53
                               Met Leu Val Met Tyr Leu Leu Ala Ala Leu Phe
                               1           5           10
    ggt tac cta acc ttc tat gga gaa gtt gaa gat gaa tta ctt cat gcc      101
    Gly Tyr Leu Thr Phe Tyr Gly Glu Val Glu Asp Glu Leu Leu His Ala
10  tac agc aaa gtg tat aca tta gac atc cct ctt ctc atg gtt cgc ctg      149
    Tyr Ser Lys Val Tyr Thr Leu Asp Ile Pro Leu Leu Met Val Arg Leu
                               15           20           25           30           35           40
    gca gtc ctt gtg gca gta aca cta act gtg ccc att gtc ctc ttc cca      197
15  Ala Val Leu Val Ala Val Thr Leu Thr Val Pro Ile Val Leu Phe Pro
    att cgt aca tca gtg atc aca ctg tta ttt ccc aaa cga ccc ttc agc      245
    Ile Arg Thr Ser Val Ile Thr Leu Leu Phe Pro Lys Arg Pro Phe Ser
    60                               65                               70                               75
20  tgg ata cga cat ttc ctg att gca gct gtg ctt att gca ctt aat aat      293
    Trp Ile Arg His Phe Leu Ile Ala Ala Val Leu Ile Ala Leu Asn Asn
                               80                               85                               90
    gtt ctg gtc atc ctt gtg cca act ata aaa tac atc ttc gga ttc ata      341
    Val Leu Val Ile Leu Val Pro Thr Ile Lys Tyr Ile Phe Gly Phe Ile
25  ggg gct tct tct gcc act atg ctg att ttt att ctt cca gca gtt ttt      389
    Gly Ala Ser Ser Ala Thr Met Leu Ile Phe Ile Leu Pro Ala Val Phe
    110                               115                               120
    tat ctt aaa ctt gtc aag aaa gaa act ttt agg tca ccc caa aag gtc      437
30  Tyr Leu Lys Leu Val Lys Lys Glu Thr Phe Arg Ser Pro Gln Lys Val
    ggg gct tta att ttc ctt gtg gtt gga ata ttc ttc atg att gga agc      485
    Gly Ala Leu Ile Phe Leu Val Val Gly Ile Phe Phe Met Ile Gly Ser
    140                               145                               150                               155
35  atg gca ctc att ata att gac tgg att tat gat cct cca aat tcc aag      533
    Met Ala Leu Ile Ile Ile Asp Trp Ile Tyr Asp Pro Pro Asn Ser Lys
    160                               165                               170
    cat cac taacacaagg aaaaatactt tctttttcta ttggaaatgg ttacaagtta      589
    His His
40  tactccaaaa gatatttgaa ttatcttgat tggaatgtta ttcataggaa ataacaggaa      649
    gattccaaag acgtttacca gtmatatcac caggcacctg cagaagagga aaatcactgt      709
    ttttgtcaag gatggttggtg tatgtgttta aaataaaaacc tgtggtgcac aaaaaaaaaa      769
    aaaaaaa      776

```

45 <210> 215

<211> 1412

<212> DNA

<213> Homo sapiens

50 <220>

<221> CDS

<222> 34..1143

<400> 215

```

55  atgcggtgaa gggcgagcgg cgcggcggct gcg atg agt gcc tct gcg gcc acc      54
                               Met Ser Ala Ser Ala Ala Thr
                               1           5
    ggg gtc ttc gtg ctg tcc ctc tcg gcc atc ccg gtc acc tat gtc ttc      102
    Gly Val Phe Val Leu Ser Leu Ser Ala Ile Pro Val Thr Tyr Val Phe
60  aac cac ctg gcg gcc cag cat gat tcc tgg act att gta ggg gtt gct      150
    Asn His Leu Ala Ala Gln His Asp Ser Trp Thr Ile Val Gly Val Ala
    25           30           35

```

	gcc	ctc	atc	ctg	ttc	ctg	gta	gca	ctg	ctg	gct	cgt	gtc	ctc	gtc	aaa	198
	Ala	Leu	Ile	Leu	Phe	Leu	Val	Ala	Leu	Leu	Ala	Arg	Val	Leu	Val	Lys	
	40					45					50					55	
	aga	aaa	cca	ccc	cgg	gac	cca	ctg	ttc	tat	gtg	tat	gca	gtt	ttt	gga	246
5	Arg	Lys	Pro	Pro	Arg	Asp	Pro	Leu	Phe	Tyr	Val	Tyr	Ala	Val	Phe	Gly	
					60					65					70		
	ttt	acc	agc	gtg	gtg	aac	ctc	atc	ata	gga	ctg	gag	caa	gat	gga	atc	294
	Phe	Thr	Ser	Val	Val	Asn	Leu	Ile	Ile	Gly	Leu	Glu	Gln	Asp	Gly	Ile	
				75					80					85			
10	att	gac	ggg	ttc	atg	aca	cac	tac	ttg	aga	gag	ggg	gaa	ccg	tat	ctg	342
	Ile	Asp	Gly	Phe	Met	Thr	His	Tyr	Leu	Arg	Glu	Gly	Glu	Pro	Tyr	Leu	
			90					95					100				
	aac	acc	gca	tat	ggg	cac	atg	atc	tgc	tac	tgg	gat	ggc	tct	gct	cat	390
	Asn	Thr	Ala	Tyr	Gly	His	Met	Ile	Cys	Tyr	Trp	Asp	Gly	Ser	Ala	His	
15			105				110					115					
	tat	ctg	atg	tac	ctg	gtg	atg	gtg	gca	gcc	ata	gca	tgg	gag	gaa	act	438
	Tyr	Leu	Met	Tyr	Leu	Val	Met	Val	Ala	Ala	Ile	Ala	Trp	Glu	Glu	Thr	
	120					125					130					135	
	tat	aga	acc	att	ggc	cta	tat	tgg	gtt	gga	tct	att	att	atg	agt	gtt	486
20	Tyr	Arg	Thr	Ile	Gly	Leu	Tyr	Trp	Val	Gly	Ser	Ile	Ile	Met	Ser	Val	
				140						145					150		
	gtt	gtt	ttt	gtg	cca	gga	aac	att	gta	ggg	aag	tat	gga	aca	cga	att	534
	Val	Val	Phe	Val	Pro	Gly	Asn	Ile	Val	Gly	Lys	Tyr	Gly	Thr	Arg	Ile	
				155					160					165			
25	tgc	cct	gct	ttt	ttc	tta	agc	ata	cca	tat	act	tgt	ctt	cct	gtc	tgg	582
	Cys	Pro	Ala	Phe	Phe	Leu	Ser	Ile	Pro	Tyr	Thr	Cys	Leu	Pro	Val	Trp	
			170					175					180				
	gct	ggg	ttc	aga	atc	tat	aat	cag	cca	tca	gaa	aat	tat	aat	tac	ccc	630
	Ala	Gly	Phe	Arg	Ile	Tyr	Asn	Gln	Pro	Ser	Glu	Asn	Tyr	Asn	Tyr	Pro	
30			185				190					195					
	tca	aag	gtt	att	caa	gaa	gcc	caa	gcg	aaa	gac	ctg	ctg	aga	aga	cca	678
	Ser	Lys	Val	Ile	Gln	Glu	Ala	Gln	Ala	Lys	Asp	Leu	Leu	Arg	Arg	Pro	
	200					205					210					215	
	ttt	gat	tta	atg	ttg	gtt	gtg	tgt	ctc	ctc	ctg	gca	act	gga	ttt	tgc	726
35	Phe	Asp	Leu	Met	Leu	Val	Val	Cys	Leu	Leu	Ala	Thr	Gly	Phe	Cys		
				220						225				230			
	ctg	ttc	aga	ggg	ttg	att	gct	ttg	gat	tgc	cca	tct	gag	ctc	tgc	cga	774
	Leu	Phe	Arg	Gly	Leu	Ile	Ala	Leu	Asp	Cys	Pro	Ser	Glu	Leu	Cys	Arg	
				235					240					245			
40	tta	tat	acg	caa	ttt	caa	gag	ccc	tat	cta	aag	gat	cct	gct	gct	tat	822
	Leu	Tyr	Thr	Gln	Phe	Gln	Glu	Pro	Tyr	Leu	Lys	Asp	Pro	Ala	Ala	Tyr	
			250					255					260				
	cct	aaa	att	cag	atg	ctg	gca	tat	atg	ttc	tat	tct	gtt	cct	tac	ttt	870
	Pro	Lys	Ile	Gln	Met	Leu	Ala	Tyr	Met	Phe	Tyr	Ser	Val	Pro	Tyr	Phe	
45			265				270					275					
	gtg	act	gca	ctg	tat	ggc	tta	gtg	gtt	cct	gga	tgt	tcc	tgg	atg	cct	918
	Val	Thr	Ala	Leu	Tyr	Gly	Leu	Val	Val	Pro	Gly	Cys	Ser	Trp	Met	Pro	
	280					285					290					295	
	gac	atc	aca	ttg	ata	cat	gct	gga	ggg	ctg	gct	cag	gct	cag	ttt	tct	966
50	Asp	Ile	Thr	Leu	Ile	His	Ala	Gly	Gly	Leu	Ala	Gln	Ala	Gln	Phe	Ser	
				300						305					310		
	cac	att	ggg	gca	tct	ctt	cat	gct	aga	act	gct	tat	gtc	tac	aga	gtc	1014
	His	Ile	Gly	Ala	Ser	Leu	His	Ala	Arg	Thr	Ala	Tyr	Val	Tyr	Arg	Val	
				315					320					325			
55	cct	gaa	gaa	gca	aaa	atc	ctt	ttt	tta	gca	tta	aac	ata	gca	tat	gga	1062
	Pro	Glu	Glu	Ala	Lys	Ile	Leu	Phe	Leu	Ala	Leu	Asn	Ile	Ala	Tyr	Gly	
				330				335					340				
	gtt	ctt	cct	cag	ctc	ttg	gcc	tat	cgt	tgt	atc	tac	aaa	cca	gag	ttc	1110
	Val	Leu	Pro	Gln	Leu	Leu	Ala	Tyr	Arg	Cys	Ile	Tyr	Lys	Pro	Glu	Phe	
60			345				350					355					
	ttc	ata	aaa	aca	aag	gca	gaa	gaa	aaa	gtg	gaa	taaaaaatatt	acttcatggtt				1163
	Phe	Ile	Lys	Thr	Lys	Ala	Glu	Glu	Lys	Val	Glu						
	360					365					370						

cctcctttct aaattactaa cttttgttat actggtactg atattttgtc ccatttcact 1223
 ctcttctcat acgtgagtac ttaagaatat gtacattctt gctctgcact gatatgtgtga 1283
 gctatatggt attgtgtaaa ttttttttga aggaaaatgg aaattcttga gaaacagttt 1343
 gtttaaagaa atatattcaa aatcatttgt gaataaactt gatcatccat ctcaaaaaaa 1403
 5 aaaaaaaaaa 1412

<210> 216

<211> 1773

<212> DNA

10 <213> Homo sapiens

<220>

<221> CDS

<222> 6..1184

15

<400> 216

ccaac atg acc tac agg tgg ggg aca ctg ctc atg aag aga aag ttt gag 50
 Met Thr Tyr Arg Trp Gly Thr Leu Leu Met Lys Arg Lys Phe Glu
 1 5 10 15
 20 gag ccc cgg cca gga ttt cat ggt gtc ttg ggt atc aat tcc atc act 98
 Glu Pro Arg Pro Gly Phe His Gly Val Leu Gly Ile Asn Ser Ile Thr
 20 25 30
 ggg aag gag gag cct ctg tac ccc agc tac aag aga cag ttg cgc att 146
 Gly Lys Glu Glu Pro Leu Tyr Pro Ser Tyr Lys Arg Gln Leu Arg Ile
 25 35 40 45
 tac ctg gtc tcc ctg cca ttc gtg tgc ctc tgc ctc tat ttc tca ctg 194
 Tyr Leu Val Ser Leu Pro Phe Val Cys Leu Cys Leu Tyr Phe Ser Leu
 50 55 60
 tat gtc atg atg att tac ttc gac atg gag gtt tgg gcc ttg ggt cta 242
 Tyr Val Met Met Ile Tyr Phe Asp Met Glu Val Trp Ala Leu Gly Leu
 65 70 75
 cat gag aac agc ggg tct gag tgg acc agt gtc ctg ttg tat gtg ccc 290
 His Glu Asn Ser Gly Ser Glu Trp Thr Ser Val Leu Leu Tyr Val Pro
 80 85 90 95
 35 agc atc atc tat gcc att gtg att gag atc atg aat cgt ctc tat cga 338
 Ser Ile Ile Tyr Ala Ile Val Ile Glu Ile Met Asn Arg Leu Tyr Arg
 100 105 110
 tat gct gcc gag ttt tta act tca tgg gag aat cac aga ttg gaa tct 386
 Tyr Ala Ala Glu Phe Leu Thr Ser Trp Glu Asn His Arg Leu Glu Ser
 115 120 125
 gcc tat cag aac cat cta att ctg aaa gtt tta gtg ttc aac ttc ctc 434
 Ala Tyr Gln Asn His Leu Ile Leu Lys Val Leu Val Phe Asn Phe Leu
 130 135 140
 aat tgc ttt gcc tca ctc ttc tat att gcc ttt gtc ttg aaa gat atg 482
 Asn Cys Phe Ala Ser Leu Phe Tyr Ile Ala Phe Val Leu Lys Asp Met
 145 150 155
 aag ctt ttg cgc cag agc ttg gcc act ctc cta att acc tcc cag atc 530
 Lys Leu Leu Arg Gln Ser Leu Ala Thr Leu Leu Ile Thr Ser Gln Ile
 160 165 170 175
 50 ctc aac caa att atg gaa tct ttt ctt cct tat tgg ctc caa agg aag 578
 Leu Asn Gln Ile Met Glu Ser Phe Leu Pro Tyr Trp Leu Gln Arg Lys
 180 185 190
 cat ggt gtg cgg gtg aag agg aag gtg cag gct tta aag gca gac att 626
 His Gly Val Arg Val Lys Arg Lys Val Gln Ala Leu Lys Ala Asp Ile
 195 200 205
 55 gat gct aca tta tat gaa caa gtc atc ctg gaa aaa gaa atg gga act 674
 Asp Ala Thr Leu Tyr Glu Gln Val Ile Leu Glu Lys Glu Met Gly Thr
 210 215 220
 tat ttg ggc acc ttt gat gat tac ttg gag tta ttc ctg cag ttt ggt 722
 Tyr Leu Gly Thr Phe Asp Asp Tyr Leu Glu Leu Phe Leu Gln Phe Gly
 225 230 235
 tat gtg agc ctt ttc tcc tgt gtt tac cca tta gca gct gcc ttt gct 770
 Tyr Val Ser Leu Phe Ser Cys Val Tyr Pro Leu Ala Ala Ala Phe Ala

	240		245		250		255	
	gtg tta aat aac ttc act gaa gta aat tca gat gcc tta aaa atg tgc							818
	Val Leu Asn Asn Phe Thr Glu Val Asn Ser Asp Ala Leu Lys Met Cys							
		260		265		270		
5	agg gtc ttc aaa cgt cca ttc tca gaa cct tca gcc aat att ggt gtg							866
	Arg Val Phe Lys Arg Pro Phe Ser Glu Pro Ser Ala Asn Ile Gly Val							
		275		280		285		
	tgg cag ttg gct ttt gaa acg atg agt gtt ata tct gtg gtc act aac							914
	Trp Gln Leu Ala Phe Glu Thr Met Ser Val Ile Ser Val Val Thr Asn							
10		290		295		300		
	tgt gcg ctg att gga atg tca cca caa gtg aat gca gtc ttt cca gaa							962
	Cys Ala Leu Ile Gly Met Ser Pro Gln Val Asn Ala Val Phe Pro Glu							
		305		310		315		
	tca aaa gca gac ctc att ttg att gta gta gca gtg gag cac gca ctc							1010
15	Ser Lys Ala Asp Leu Ile Leu Ile Val Val Ala Val Glu His Ala Leu							
		320		325		330		335
	ctg gct tta aag ttt ata ctt gca ttt gcc ata cct gat aag cca cgg							1058
	Leu Ala Leu Lys Phe Ile Leu Ala Phe Ala Ile Pro Asp Lys Pro Arg							
		340		345		350		
20	cat atc cag atg aaa cta gcc aga ctg gaa ttt gag tct ttg gag gca							1106
	His Ile Gln Met Lys Leu Ala Arg Leu Glu Phe Glu Ser Leu Glu Ala							
		355		360		365		
	ctc aag cag cag caa atg aag ctc gtg acc gag aac ctg aag gag gaa							1154
	Leu Lys Gln Gln Gln Met Lys Leu Val Thr Glu Asn Leu Lys Glu Glu							
25		370		375		380		
	cca atg gaa agc ggg aag gag aag gca acc tgagtgccca gcgtgccag							1204
	Pro Met Glu Ser Gly Lys Glu Lys Ala Thr							
		385		390				
	ctgccctggt ggcagaggcc tgtgtctgtg ccacacctgc cacggtggca ggggggggtac							1264
30	ccggggcagc atcgtggctc ctgaacccag acccaatgct tagccaaacg aagtggctcc							1324
	catgtggcaa gcacccttct cagtttcgca gtggcttgcc tcgggaccc tggcagttcc							1384
	cccagcccca cctgtctctc tccttcccag ttcttcccgc ggccccacac gctgctccag							1444
	ctgccaactt tgctgcagag ccaactgccgc ccttgagcct ctcaccatga gtgagccacc							1504
	agctctccac gttccccctca tagcagtgtc actcccaacc ccaccatggc ccagggaccc							1564
35	gtggacaggt tgggggatggg gtgtgtgtccc actgtgtctca tcacaggagc ctcagttgag							1624
	agtgagcggg gtacagtaag gcagtgtctc ccacactgga cctctttcct ggttctcttt							1684
	tgaatacat taacagacc tttatcaaca taacaatag taactgagct attaaaggca							1744
	aaaaaaaaa taaaaaaaaa aaaaaaaaaa							1773
40	<210> 217							
	<211> 1251							
	<212> DNA							
	<213> Homo sapiens							
45	<220>							
	<221> CDS							
	<222> 29..376							
	<400> 217							
50	tatccggtcc tcggctgcgg cgggcacc atg gtc ggt ggc gag gcg gct gcc							52
	Met Val Gly Gly Glu Ala Ala Ala							
		1		5				
	gca gtg gag gag ctg gtt tcg ggg gtg cgg cag gcg gcc gac ttc gcg							100
	Ala Val Glu Glu Leu Val Ser Gly Val Arg Gln Ala Ala Asp Phe Ala							
55		10		15		20		
	gag cag ttc cgc tcc tac tca gag agc gag aag caa tgg aag gcc cgc							148
	Glu Gln Phe Arg Ser Tyr Ser Glu Ser Glu Lys Gln Trp Lys Ala Arg							
		25		30		35		40
	atg gaa ttc atc ctg cgc cac ctg ccc gac tac cgc gac ccg ccc gac							196
60	Met Glu Phe Ile Leu Arg His Leu Pro Asp Tyr Arg Asp Pro Pro Asp							
		45		50		55		
	ggc agt ggc cgc ctg gac cag ctg ctc tcc ctc tcc atg gtc tgg gcc							244
	Gly Ser Gly Arg Leu Asp Gln Leu Leu Ser Leu Ser Met Val Trp Ala							

		60		65		70		
	aac cat ctc ttc cta ggc tgc agt tac aat aaa gac ctt tta gac aag							292
	Asn His Leu Phe Leu Gly Cys Ser Tyr Asn Lys Asp Leu Leu Asp Lys							
		75		80		85		
5	gtg atg gaa atg gcc gat ggg att gaa gtg gaa gac ctg cca caa ttt							340
	Val Met Glu Met Ala Asp Gly Ile Glu Val Glu Asp Leu Pro Gln Phe							
		90		95		100		
	act acc aga agt gaa tta atg aaa aag cat caa agc taagccagaa							386
	Thr Thr Arg Ser Glu Leu Met Lys Lys His Gln Ser							
10	105	110		115				
	gattttatcac attttcatca tcagctacag gattagaaag gaggctggga tgaatgtgac							446
	atagaccaca gcagctctct taagactcct ggtattacca acataaagag gcaggtggaa							506
	tgagaaggac tctgtctaga ttggcttttt taacattctc attttcccag gagttatcac							566
	tgtaaaagta tgcattggata tttatgtatt tataaatcat gcactctaag atgagttcat							626
15	caacattgta aaagccctct tttctgtttt cagggttttt tttttcttat cgacaaggtc							686
	tcactctgtc gccaggcag aatcacaaag gtgcattatt ggctcattgc agcctcgaac							746
	tcctgggctc atattttcag gggtttttgt tttttgtttt gtttttttga gacagagtct							806
	tgctctgttg cccaggcagt agtgcmagtg gcgcgatata ttttcagttt ttaaactca							866
	gaatttttgt ttaaaatgcc tttttgggct gggccacagt ggccttatgc ccataataat							926
20	cccagcactt tgggaggccg aggtgagcag atcacctgag gttaggagtt tgagaccagc							986
	ctggccaaca cgatgaaacc ccgtctctac taaaaatata aacaaaatta gctgggcatg							1046
	gtggcggaaca tctgtaatcc cagctactca ggaggctgaa gcagaagaac tgcttgaacc							1106
	tgggaggtgg aggttgcagt gagccaagat cgcaccattg cactccatcc tgggcgacaa							1166
	aatgaaaca ccgtctcaaa aaaaaataaa aaataataaa ataaaatgcc tttttgttgt							1226
25	tgatgtgaaa aaaaaaaaaa aaaaa							1251
	<210> 218							
	<211> 894							
	<212> DNA							
30	<213> Homo sapiens							
	<220>							
	<221> CDS							
	<222> 78..566							
35	<400> 218							
	gcgcgccatc ttggctccgg atcgtgcgtg aggcggcttc gtgggcagcg agagtcacag							60
	acaagacagc aagcagg atg gag cac tac cgg aaa gct ggc tct gta gag							110
	Met Glu His Tyr Arg Lys Ala Gly Ser Val Glu							
40		1		5		10		
	ctc cca gcg cct tcc cca atg ccc cag cta cct cct gat acc ctt gag							158
	Leu Pro Ala Pro Ser Pro Met Pro Gln Leu Pro Pro Asp Thr Leu Glu							
		15		20		25		
	atg cgg gtc cga gat ggc agc aaa att cgc aac ctg ctg ggg ttg gct							206
45	Met Arg Val Arg Asp Gly Ser Lys Ile Arg Asn Leu Leu Gly Leu Ala							
		30		35		40		
	ctg ggt cgg ttg gag ggc ggc agt gct cgg cat gta gtg ttc tca ggt							254
	Leu Gly Arg Leu Glu Gly Gly Ser Ala Arg His Val Val Phe Ser Gly							
		45		50		55		
50	tct ggc agg gct gca gga aag gct gtc agc tgc gct gag att gtc aag							302
	Ser Gly Arg Ala Ala Gly Lys Ala Val Ser Cys Ala Glu Ile Val Lys							
		60		65		70		75
	cgg cgg gtc cca ggc ctg cac cag ctc acc aag cta cgt ttc ctt cag							350
	Arg Arg Val Pro Gly Leu His Gln Leu Thr Lys Leu Arg Phe Leu Gln							
55		80		85		90		
	act gag gac agc tgg gtc cca gcc tca cct gac aca ggg cta gac ccc							398
	Thr Glu Asp Ser Trp Val Pro Ala Ser Pro Asp Thr Gly Leu Asp Pro							
		95		100		105		
	ctc aca gtg cgc cgc cat gtg cct gca gtg tgg gtg ctg ctc agc cgg							446
60	Leu Thr Val Arg Arg His Val Pro Ala Val Trp Val Leu Leu Ser Arg							
		110		115		120		
	gac ccc ctg gac ccc aat gag tgt ggt tac caa ccc cca gga gca ccc							494
	Asp Pro Leu Asp Pro Asn Glu Cys Gly Tyr Gln Pro Pro Gly Ala Pro							

	125	130	135	
	cct ggc ctg ggt tcc atg ccc agc tcc agc tgt ggc cct cgt tcc cga			542
	Pro Gly Leu Gly Ser Met Pro Ser Ser Ser Cys Gly Pro Arg Ser Arg			
	140	145	150	155
5	aga agg gct cga gac acc cga tgc tgaagacctg ctgagccagc ctgttctccg			596
	Arg Arg Ala Arg Asp Thr Arg Ser			
	160			
	ggcctgaatg tctgggggtgc ttgtgccttt tctgagaagc gttgtgactg ctcaacatcc			656
	ccatcaagggt ttgagtccac aaaagtggac ctccctatca tgcttcccct tccctctagc			716
10	atgtgggaag ggactgctgt gaagaatgac agatgtgggg cctctgccaa gttctgcatt			776
	gctaaataag ggcttctctt gccttctacc tacagtgcatt ttgaactgcc ttctgaaaga			836
	ggtccaggga gggatttagg aaataaagtt tctacctatt taaaaaaaaa aaaaaaaaaa			894
	<210> 219			
15	<211> 910			
	<212> DNA			
	<213> Homo sapiens			
	<220>			
20	<221> CDS			
	<222> 16..705			
	<400> 219			
25	acatgagcca ccaaaa atg gtg gtg ttc ggg tat gag gct ggg act aag cca			51
	Met Val Val Phe Gly Tyr Glu Ala Gly Thr Lys Pro			
	1 5 10			
	agg gat tca ggt gtg gtg ccg gtg gga act gag gaa gcg ccc aag gtt			99
	Arg Asp Ser Gly Val Val Pro Val Gly Thr Glu Glu Ala Pro Lys Val			
	15 20 25			
30	ttc aag atg gca gca tct atg cat ggt cag ccc agt cct tct cta gaa			147
	Phe Lys Met Ala Ala Ser Met His Gly Gln Pro Ser Pro Ser Leu Glu			
	30 35 40			
	gat gca aaa ctc aga aga cca atg gtc ata gaa atc ata gaa aaa aat			195
	Asp Ala Lys Leu Arg Arg Pro Met Val Ile Glu Ile Ile Glu Lys Asn			
35	45 50 55 60			
	ttt gac tat ctt aga aaa gaa atg aca caa aat ata tat caa atg gcg			243
	Phe Asp Tyr Leu Arg Lys Glu Met Thr Gln Asn Ile Tyr Gln Met Ala			
	65 70 75			
	aca ttt gga aca aca gct ggt ttc tct gga ata ttc tca aac ttc ctg			291
40	Thr Phe Gly Thr Thr Ala Gly Phe Ser Gly Ile Phe Ser Asn Phe Leu			
	80 85 90			
	ttc aga cgc tgc ttc aag gtt aaa cat gat gct ttg aag aca tat gca			339
	Phe Arg Arg Cys Phe Lys Val Lys His Asp Ala Leu Lys Thr Tyr Ala			
	95 100 105			
45	tca ttg gct aca ctt cca ttt ttg tct act gtt gtt act gac aag ctt			387
	Ser Leu Ala Thr Leu Pro Phe Leu Ser Thr Val Val Thr Asp Lys Leu			
	110 115 120			
	ttt gta att gat gct ttg tat tca gat aat ata agc aag gaa aac tgt			435
	Phe Val Ile Asp Ala Leu Tyr Ser Asp Asn Ile Ser Lys Glu Asn Cys			
50	125 130 135 140			
	gtt ttc aga agc tca ctg att ggc ata gtt tgt ggt gtt ttc tat ccc			483
	Val Phe Arg Ser Ser Leu Ile Gly Ile Val Cys Gly Val Phe Tyr Pro			
	145 150 155			
	agt tct ttg gct ttt act aaa aat gga cgc ctg gca acc aag tat cat			531
55	Ser Ser Leu Ala Phe Thr Lys Asn Gly Arg Leu Ala Thr Lys Tyr His			
	160 165 170			
	acc gtt cca ctg cca cca aaa gga agg gtt tta atc cat tgg atg acg			579
	Thr Val Pro Leu Pro Pro Lys Gly Arg Val Leu Ile His Trp Met Thr			
	175 180 185			
60	ctt tgt caa aca caa atg aaa tta atg gcg att cct cta gtc ttt cag			627
	Leu Cys Gln Thr Gln Met Lys Leu Met Ala Ile Pro Leu Val Phe Gln			
	190 195 200			
	att atg ttt gga ata tta aat ggt cta tac cat tat gca gta ttt gaa			675

Ile Met Phe Gly Ile Leu Asn Gly Leu Tyr His Tyr Ala Val Phe Glu
 205 210 215 220
 gag aca ctt gag aaa act ata cat gaa gag taaccaaaaa aatgaatggt 725
 Glu Thr Leu Glu Lys Thr Ile His Glu Glu
 5 225 230
 tgctaactta gcaaaatgaa gtttctataa agaggactca ggcattgctg aaagagttaa 785
 aagtaactgt gaacaaataa tttgttctgt gccttttgcc tggtatatag caaataactca 845
 aaaaatattc aataattcaa tcaataaata taagtttcat cttacaccaa aaaaaaaaaa 905
 aaaaa 910
 10
 <210> 220
 <211> 519
 <212> DNA
 <213> Homo sapiens
 15
 <220>
 <221> CDS
 <222> 103..405
 20 <400> 220
 acttccggtg cgaaccgcct cggccgttcc ctgcgaggagc ttactgagcg cggccgcgga 60
 gcccagctcc gccgccgagc gcctgtgccg gcacgbhaca cc atg gag cgc ccg 114
 Met Glu Arg Pro
 1
 25 gat aag gcg gcg ctg aac gca ctg cag cct cct gag ttc aga aat gaa 162
 Asp Lys Ala Ala Leu Asn Ala Leu Gln Pro Pro Glu Phe Arg Asn Glu
 5 10 15 20
 agc tca tta gca tct aca ctg aag acg ctc ctg ttc ttc aca gct tta 210
 Ser Ser Leu Ala Ser Thr Leu Lys Thr Leu Leu Phe Phe Thr Ala Leu
 30 25 30 35
 atg atc act gtt cct att ggg tta tat ttc aca act aaa tct tac ata 258
 Met Ile Thr Val Pro Ile Gly Leu Tyr Phe Thr Thr Lys Ser Tyr Ile
 40 45 50
 ttt gaa ggc gcc ctt ggg atg tcc aat agg gac agc tat ttt tac gct 306
 35 Phe Glu Gly Ala Leu Gly Met Ser Asn Arg Asp Ser Tyr Phe Tyr Ala
 55 60 65
 gct att gtt gca gtg gtc gcc gtc cat gtg gtg ctg gcc ctc ttt gtg 354
 Ala Ile Val Ala Val Val Ala Val His Val Val Leu Ala Leu Phe Val
 70 75 80
 40 tat gtg gcc tgg aat gaa ggc tca cga cag tgb cgt gaa ggc aaa cag 402
 Tyr Val Ala Trp Asn Glu Gly Ser Arg Gln Xaa Arg Glu Gly Lys Gln
 85 90 95 100
 gat taaagtgaac atcacctttt tatagcatta aattcatttt ttaaaatgat 455
 Asp
 45 aatgctggag ggggccatct gatttgaata aagttgaaag aacatgtaaa aaaaaaaaaa 515
 aaaa 519
 <210> 221
 <211> 632
 50 <212> DNA
 <213> Homo sapiens
 <220>
 <221> CDS
 55 <222> 72..350
 <400> 221
 agtgagaccg cgcggcaaca gcttgccggt gcggtagtcc cgtgggagcgt ccgctggctg 60
 tgcaggcggc c atg gat tcc ttg cgg aaa atg ctg atc tca gtc gca atg 110
 60 Met Asp Ser Leu Arg Lys Met Leu Ile Ser Val Ala Met
 1 5 10
 ctg ggc gca ggg gct ggc gtg ggc tac gcg ctc ctc gtt atc gtg acc 158
 Leu Gly Ala Gly Ala Gly Val Gly Tyr Ala Leu Leu Val Ile Val Thr

220

<212> DNA

<213> Homo sapiens

<220>

5 <221> CDS

<222> 38...322

<400> 223

```

10 actgctgtcc cccgagctgc tctacgcgct ggcgcgg atg ggg cac ggg gac gag      55
                                   Met Gly His Gly Asp Glu
                                   1           5
   atc gtt ctt gcg gac ttg aac ttc ccg gcc tcc tcc atc tgc cag tgt      103
   Ile Val Leu Ala Asp Leu Asn Phe Pro Ala Ser Ser Ile Cys Gln Cys
               10           15           20
15 ggg ccc atg gag atc cgt gca gac ggc ctg ggc atc ccg cag ctc ctg      151
   Gly Pro Met Glu Ile Arg Ala Asp Gly Leu Gly Ile Pro Gln Leu Leu
               25           30           35
   gag gcc gtg cta gct gct gcc cct gga cac cta tgt gga gag tcc ggc      199
   Glu Ala Val Leu Ala Ala Pro Gly His Leu Cys Gly Glu Ser Gly
20           40           45           50
   tgc agt cat gga gct ggt gcc cag cga caa gga gag ggg cct gca gac      247
   Cys Ser His Gly Ala Gly Ala Gln Arg Gln Gly Glu Gly Pro Ala Asp
   55           60           65           70
25 ccc agt gtg gac gga gta cga gtc cat cct acg cag ggc cgg ctg tgt      295
   Pro Ser Val Asp Gly Val Arg Val His Pro Thr Gln Gly Arg Leu Cys
               75           80           85
   gag agc cct ggc aaa gat aga gag gtt tgagttttat gaacgggcta      342
   Glu Ser Pro Gly Lys Asp Arg Glu Val
               90           95
30 agaaggcttt tgctgttggtg gcaacggggg agacggccct ctacggaaac ctcatcctca      402
   ggaagggggt gcttgccctc aacccctgc ttagggcctg gtgaagacca cctgggcccgg      462
   aagaggaact gggggcacc ctagctccag taccaccact cacaacaggc ctcccagtgg      522
   cagctcccag acctggggcc tggccagggc tctaggggcc ggcagtcttg ggggtgggcc      582
   tgccaattgg gacgagtatc cctgatttgt gaaaatgatg gaaaaacgtt caaaaaaaaa      642
35 aaaaaaaaaa                                     650

```

<210> 224

<211> 502

<212> DNA

40 <213> Homo sapiens

<220>

<221> CDS

<222> 202...480

45

<400> 224

```

   attctaaggc tacaggtoct ttggcaactg ctccccctct tccttctcct tttttttcaa      60
   tctaatatgt gggacatacc aagtgatgag atcctttctg gcatacctaag aaaatgccaa      120
   ctatagaaca acagatgtag tctccttatt aagtctgaag accaaacttc ttagtgcaaa      180
50 gcagtcaagt cttttctcaa c atg acc cca atc aag ctt ttg aac tta aca      231
                                   Met Thr Pro Ile Lys Leu Leu Asn Leu Thr
                                   1           5           10
   tca aga tat aac ttc aga aga acg ttt gga ata gag ctc agt tca aac      279
   Ser Arg Tyr Asn Phe Arg Arg Thr Phe Gly Ile Glu Leu Ser Ser Asn
               15           20           25
55 tct tcc tat tgc aaa cga gga aat ggc tac aga agc aga gtg ccc aaa      327
   Ser Ser Tyr Cys Lys Arg Gly Asn Gly Tyr Arg Ser Arg Val Pro Lys
               30           35           40
   gaa tgc gaa tgc aac tgg ctt cat ctt gaa agc gac act ctg aag aaa      375
60 Glu Cys Glu Cys Asn Trp Leu His Leu Glu Ser Asp Thr Leu Lys Lys
               45           50           55
   tta ccc ata att tct ccc tct tgg aca tgc aga att atc ctg ttc ttg      423
   Leu Pro Ile Ile Ser Pro Ser Trp Thr Cys Arg Ile Ile Leu Phe Leu

```

```

      60      65      70
tat ttt tct ggc cag ctt ctc caa ctt tcc ctt tct tgt ttg caa cta      471
Tyr Phe Ser Gly Gln Leu Leu Gln Leu Ser Leu Ser Cys Leu Gln Leu
75      80      85      90
5 att aaa ctt taaggataaa aaaaaaaaaa aa      502
Ile Lys Leu

<210> 225
<211> 1739
10 <212> DNA
    <213> Homo sapiens

    <220>
    <221> CDS
15 <222> 171..1670

    <400> 225
actctggcct tgctgcttct ctccagctcc tgaacttttc tttcttccat catgctctga      60
gccattcct tgaaaactaa aagggtccctg actcccagtc tgcagccatc ctgggcctgc      120
20 tgagctctga ttcaagtgcc tgcctctgcc ccttggtggg ctgaagcttc atg gag      176
    Met Glu
    1
gta tcc acc aac ccc tcc tcc aac atc gat cca ggc aac tat gtt gaa      224
Val Ser Thr Asn Pro Ser Ser Asn Ile Asp Pro Gly Asn Tyr Val Glu
25      5      10      15
atg aat gat tca atc acc cac cta ccc tct aaa gtg gtg ata caa gat      272
Met Asn Asp Ser Ile Thr His Leu Pro Ser Lys Val Val Ile Gln Asp
20      25      30
att act atg gag cta cac tgc cct ctg tgc aat gat tgg ttc cga gac      320
Ile Thr Met Glu Leu His Cys Pro Leu Cys Asn Asp Trp Phe Arg Asp
35      40      45      50
cca ctg atg cta agc tgt ggc cac aac ttc tgt gaa gcc tgt atc caa      368
Pro Leu Met Leu Ser Cys Gly His Asn Phe Cys Glu Ala Cys Ile Gln
55      60      65
35 gac ttt tgg agg ctg caa gca aag gaa aca ttc tgt cct gag tgt aag      416
Asp Phe Trp Arg Leu Gln Ala Lys Glu Thr Phe Cys Pro Glu Cys Lys
70      75      80
atg cta tgt cag tat aac aac tgt aca ttc aac cct gta ctg gac aag      464
Met Leu Cys Gln Tyr Asn Asn Cys Thr Phe Asn Pro Val Leu Asp Lys
85      90      95
40 ttg gta gag aag att aag aag tta ccc tta ctc aag ggc cat cca cag      512
Leu Val Glu Lys Ile Lys Lys Leu Pro Leu Leu Lys Gly His Pro Gln
100      105      110
tgc cca gag cat gga gag aac ctg aaa ctg ttc agt aaa cca gat ggg      560
Cys Pro Glu His Gly Glu Asn Leu Lys Leu Phe Ser Lys Pro Asp Gly
115      120      125      130
aaa ctg atc tgc ttt caa tgc aag gat gct cgg ttg tct gtg ggg cag      608
Lys Leu Ile Cys Phe Gln Cys Lys Asp Ala Arg Leu Ser Val Gly Gln
135      140      145
50 tct aag gag ttc ctg caa atc tct gat gct gtc cat ttc ttc atg gag      656
Ser Lys Glu Phe Leu Gln Ile Ser Asp Ala Val His Phe Phe Met Glu
150      155      160
gag ctt gcc atc caa cag ggt caa ctg gag aca act ctg aag gag ctt      704
Glu Leu Ala Ile Gln Gln Gly Gln Leu Glu Thr Thr Leu Lys Glu Leu
165      170      175
55 cag acc ctg agg aac atg cag aag gaa gct att gct gct cac aag gaa      752
Gln Thr Leu Arg Asn Met Gln Lys Glu Ala Ile Ala Ala His Lys Glu
180      185      190
aac aag cta cat ctg cag caa cat gtg tcc atg gag ttt cta aag ctg      800
Asn Lys Leu His Leu Gln Gln His Val Ser Met Glu Phe Leu Lys Leu
195      200      205      210
cat cag ttc ctg cac agc aaa gaa aag gac att tta act gag ctc cgg      848
His Gln Phe Leu His Ser Lys Glu Lys Asp Ile Leu Thr Glu Leu Arg

```

223

<221> CDS

<222> 199..618

<400> 226

```

5 aactggatag agtactgccc ccttcagccc atggagaaaag gcaaatgcct ccttcagagt      60
  ctacctaatag ctttctcaga taaataagca tgaagaaaag tcaaagtcca ttctagctct      120
  aaaataagga atgaaatggt ttcctgatat gattttttgt ttctatctga taataatttt      180
  atatatcaca gaaacagc atg gtt ctt act aaa cct ctt caa aga aat ggc      231
                        Met Val Leu Thr Lys Pro Leu Gln Arg Asn Gly
10                               1                               5                               10
  agc atg atg agc ttt gaa aat gtg aaa gaa aag agc aga gaa gga ggg      279
  Ser Met Met Ser Phe Glu Asn Val Lys Glu Lys Ser Arg Glu Gly Gly
                        15                               20                               25
  ccc cat gca cac aca ccc gaa gaa gaa ttg tgt ttc gtg gta aca cac      327
15 Pro His Ala His Thr Pro Glu Glu Glu Leu Cys Phe Val Val Thr His
                        30                               35                               40
  tac cct cag gtt cag acc aca ctc aac ctg ttt ttc cat ata ttc aag      375
  Tyr Pro Gln Val Gln Thr Thr Leu Asn Leu Phe Phe His Ile Phe Lys
                        45                               50                               55
20 gtt ctt act caa cca ctt tcc ctt ctg tgg ggt tgt gat cag aag cct      423
  Val Leu Thr Gln Pro Leu Ser Leu Leu Trp Gly Cys Asp Gln Lys Pro
  60                               65                               70                               75
  cgt act gtt cct acc ctt gga aac ggc gca tgg gat acc tgc caa caa      471
  Arg Thr Val Pro Thr Leu Gly Asn Gly Ala Trp Asp Thr Cys Gln Gln
25                               80                               85                               90
  cac ata cgc act tca tca tgg aca gca aac aca ctc gtc att caa aac      519
  His Ile Arg Thr Ser Ser Trp Thr Ala Asn Thr Leu Val Ile Gln Asn
                        95                               100                               105
  cag cat tca cgg gaa agc act gtt tct gtt tgc ctt ttt atg tta atc      567
30 Gln His Ser Arg Glu Ser Thr Val Ser Val Cys Leu Phe Met Leu Ile
                        110                               115                               120
  cgc atg caa cat att ttg aaa aca gat aca ctt caa cag ttc aga ata      615
  Arg Met Gln His Ile Leu Lys Thr Asp Thr Leu Gln Gln Phe Arg Ile
                        125                               130                               135
35 tgc tagtactaat aaaaccaaca tgtaaaaaaa aaaaaaaaaa      657
  Cys
  140

```

<210> 227

40 <211> 888

<212> DNA

<213> Homo sapiens

<220>

45 <221> CDS

<222> 182..481

<400> 227

```

50 attttgccctc tcagtgttca agcttgagcc cacgcatcca actcctgaga tcttactggg      60
  aagctgctga tcatcagttt caggaagtca gcatggatca gccttacggt catggcctcc      120
  aggccttatt ctctgcctc acagggaccg gccaggatct ctatccttac agcacgttgg      180
  a atg tat atg ctc ctc tcc cca cat cgc ctt agg gag cag gca ggt gtc      229
  Met Tyr Met Leu Leu Ser Pro His Arg Leu Arg Glu Gln Ala Gly Val
  1                               5                               10                               15
55 agg ggc agc ata agg acg gcc aac agg aca gaa gac ggg ttg aag atc      277
  Arg Gly Ser Ile Arg Thr Ala Asn Arg Thr Glu Asp Gly Leu Lys Ile
                        20                               25                               30
  cga gag gct gag tca ctt cca caa agt aac aca gct gat ttt aaa tgc      325
  Arg Glu Ala Glu Ser Leu Pro Gln Ser Asn Thr Ala Asp Phe Lys Cys
60                               35                               40                               45
  ctg cat tca gca tcc ctg cag cag gct cca ggt gga att cta atg gga      373
  Leu His Ser Ala Ser Leu Gln Gln Ala Pro Gly Gly Ile Leu Met Gly
                        50                               55                               60

```

```

cca gcc tcc agt ccc tgg acc tta gcc gtg gaa gga gag aag agg aca      421
Pro Ala Ser Ser Pro Trp Thr Leu Ala Val Glu Gly Glu Lys Arg Thr
65              70              75              80
tct gca cct cct ctc aga gaa agc ctg atg cct act aaa gga ctt ggg      469
5 Ser Ala Pro Pro Leu Arg Glu Ser Leu Met Pro Thr Lys Gly Leu Gly
85              90              95
tgg tgg acg cag tgaccctcag tctggagctt gttcactgaa cattggagac      521
Trp Trp Thr Gln
100
10 tatcatttgc gcagatggtc ttgggcctct atgagcagca ggctgcaccc cacagtgacc      581
tctcattct actctgaggc atcttcatga aagcagatgt ccattgaaaa gcacccaagt      641
gcagtctcag ctgatgaact tcagaggcga ttgagacaaa ggctctcggg cccctctgcc      701
cttggatggg gcctctggta tgcacttggc ctctgtgtct ttatttagac tggtcacttc      761
acaacccatc atgtcacccc acccctaacc gtgcccactc tgggtcctcc cctcaactgc      821
15 ctgacttccc actttgagct cagcaaaggc aatagatgtt ttgtctgctt cgaaaaaaaa      881
aaaaaaa
<210> 228
<211> 716
20 <212> DNA
<213> Homo sapiens
<220>
<221> CDS
25 <222> 161..517
<400> 228
acctgtcatt atgcttacta acgttcggga cgtctcccgg gctgcttggg cgaggagagg      60
caggggtgtg tgaccccggg ggttactgtg ctgcgctaga gcacctaggg cctgctgaag      120
30 cctccctcgc ccgcgcctc tccttagtcc ttgagatgag atg gca agt tac agc      175
Met Ala Ser Tyr Ser
1              5
ggc ttc tcc ggc ctg ctg gag att cgc tac ggg cca gga cac cgc agc      223
Gly Phe Ser Gly Leu Leu Glu Ile Arg Tyr Gly Pro Gly His Arg Ser
10              15              20
tgc ctt ccc caa ttc gct ttc ttt ccg cag ccg ccg ctg ccc cga ccc      271
Cys Leu Pro Gln Phe Ala Phe Phe Pro Gln Pro Pro Leu Pro Arg Pro
25              30              35
cgg atc tgc atg tgg gtg ctg gct gag ctg ctg gag cta ggg tgt cct      319
40 Arg Ile Cys Met Trp Val Leu Ala Glu Leu Leu Glu Leu Gly Cys Pro
40              45              50
gag cag agc ctg agg gac gcc atc acc ctg gac ctc ttc tgc cac gcg      367
Glu Gln Ser Leu Arg Asp Ala Ile Thr Leu Asp Leu Phe Cys His Ala
55              60              65
ctc att ttc tgc cgc cag cag ggc ttc tca ctg gag cag acg tca gcg      415
Leu Ile Phe Cys Arg Gln Gln Gly Phe Ser Leu Glu Gln Thr Ser Ala
70              75              80              85
gct tgt gcc ctg ctc cag gat ctt cac aag gct tgt att ggt gag agg      463
Ala Cys Ala Leu Leu Gln Asp Leu His Lys Ala Cys Ile Gly Glu Arg
90              95              100
ggg cag cta cca ggt ttg agc ccc agg gag aag agg aac cgg gcc tgg      511
Gly Gln Leu Pro Gly Leu Ser Pro Arg Glu Lys Arg Asn Arg Ala Trp
105              110              115
cac aag tgaccatggg aagcagaagc aggggatttc tgcttgaat atgtcattat      567
55 His Lys
tagtagcatc atcatcacaca agccatcagc tttccaatcc actgcttccct tatctagaaa      627
ttaaggatac agcacacatt ttacaggact gttctgagaa ataatatatg caaatatatg      687
catagtgcac aataaaaaaaaa aaaaaaaaaa      716
60 <210> 229
<211> 654
<212> DNA
<213> Homo sapiens

```

<220>
 <221> CDS
 <222> 86..505
 5
 <400> 229
 agttcgcggt gtcagcctcc gcctccgagc ctcagttgtc ttctctgtga ggtgggaatg 60
 ccggtgaatc ctgccgctgg cgtgg atg aga agt gaa tgc gtg ctc gga gct 112
 Met Arg Ser Glu Cys Val Leu Gly Ala
 1
 10 gcg agt gac agc ggg cag gag gcg ccc agg gac act tgg ttt ctc cag 160
 Ala Ser Asp Ser Gly Gln Glu Ala Pro Arg Asp Thr Trp Phe Leu Gln
 10 15 20 25
 ggc tgg aag gct tct aga agg ttc ctc atc aag gga agt gtg gct ggg 208
 15 Gly Trp Lys Ala Ser Arg Arg Phe Leu Ile Lys Gly Ser Val Ala Gly
 30 35 40
 ggc gcc gtc tac ctg gtg tac gac cag gag ctg ctg ggg ccc agc gac 256
 Gly Ala Val Tyr Leu Val Tyr Asp Gln Glu Leu Leu Gly Pro Ser Asp
 45 50 55
 20 aag agc cag gca gcc cta cag aag gct ggg gag gtg gtc ccc ccc gcc 304
 Lys Ser Gln Ala Ala Leu Gln Lys Ala Gly Glu Val Val Pro Pro Ala
 60 65 70
 atg tac cag ttc agc cag tac gtg tgt cag cag aca ggc ctg cag ata 352
 Met Tyr Gln Phe Ser Gln Tyr Val Cys Gln Gln Thr Gly Leu Gln Ile
 75 80 85
 25 ccc cag ctc cca gcc cct cca aag att tac ttt ccc atc cgt gac tcc 400
 Pro Gln Leu Pro Ala Pro Pro Lys Ile Tyr Phe Pro Ile Arg Asp Ser
 90 95 100 105
 tgg aat gca ggc atc atg acg gtg atg tca gct ctg tcg gtg gcc ccc 448
 30 Trp Asn Ala Gly Ile Met Thr Val Met Ser Ala Leu Ser Val Ala Pro
 110 115 120
 tcc aag gcc cgc gag tac tcc aag gag ggc tgg gag tat gtg aag gcg 496
 Ser Lys Ala Arg Glu Tyr Ser Lys Glu Gly Trp Glu Tyr Val Lys Ala
 125 130 135
 35 cgc acc aag tagcgagtca gcaggggccc cctgccccgg ccagaacggg 545
 Arg Thr Lys
 140
 cagggctgcc actgacctga agactccgga ctgggacccc actccgaggg cagctccccg 605
 ccttgccggc ccaataaagg acttcagaag tcaaaaaaaaa aaaaaaaaaa 654
 40
 <210> 230
 <211> 635
 <212> DNA
 <213> Homo sapiens
 45
 <220>
 <221> CDS
 <222> 56..382
 50 <400> 230
 aattcgggtg gagctgagcc ggagacaggc agttgtgaaa aatttcagga caaaa atg 58
 Met
 1
 ttt cat tta agg act tgt gct gct aag ttg agg cca ttg acg gct tcc 106
 55 Phe His Leu Arg Thr Cys Ala Ala Lys Leu Arg Pro Leu Thr Ala Ser
 5 10 15
 cag act gtt aag aca ttt tca caa aac aga cca gca gca gct agg aca 154
 Gln Thr Val Lys Thr Phe Ser Gln Asn Arg Pro Ala Ala Ala Arg Thr
 20 25 30
 60 ttt caa cag att cgg tgc tat tct gca cct gtt gct gct gag ccc ttt 202
 Phe Gln Gln Ile Arg Cys Tyr Ser Ala Pro Val Ala Ala Glu Pro Phe
 35 40 45
 ctc agt ggg act agt tcg aac tat gtg gag gag atg tac tgt gct tgg 250

	Leu	Ser	Gly	Thr	Ser	Ser	Asn	Tyr	Val	Glu	Glu	Met	Tyr	Cys	Ala	Trp			
	50					55					60					65			
	ctg	gaa	aac	ccc	aaa	agt	gta	cat	aag	aca	ggg	tcc	cac	tgt	tgt	cca	298		
	Leu	Glu	Asn	Pro	Lys	Ser	Val	His	Lys	Thr	Gly	Ser	His	Cys	Cys	Pro			
5					70					75					80				
	ggc	tgg	agt	gca	gtg	gcg	gga	tct	cgg	ctt	gct	gca	acc	tcc	gac	tcc	346		
	Gly	Trp	Ser	Ala	Val	Ala	Gly	Ser	Arg	Leu	Ala	Ala	Thr	Ser	Asp	Ser			
				85					90					95					
	tgg	gtt	caa	gtg	att	ctt	atg	cct	cag	cct	ccc	gag	taact	ggg	gac		392		
10	Trp	Val	Gln	Val	Ile	Leu	Met	Pro	Gln	Pro	Pro	Glu							
		100							105										
	tac	agg	tgc	acg	cct	gact	agt	tttt	gtattt	ttag	tagaga	tggg	atttta				452		
	cttt	gtt	ggc	cagg	ctg	gtc	ttga	acc	cct	ggc	ctca	agt	gatcc	accca	cctt	ggc	ctc	512	
	ccaa	agt	gct	ggg	att	acag	gtat	gat	caa	ccac	gcct	gg	ccat	gtcat	g	cctt	gtg	aca	572
15	gaatt	ccttt	attct	gtttt	gag	cca	ataa	atatt	ttag	gttt	cga	aaa	aaaaaaaa				632		
	aaa																635		
	<210> 231																		
	<211> 634																		
20	<212> DNA																		
	<213> Homo sapiens																		
	<220>																		
	<221> CDS																		
25	<222> 56..355																		
	<400> 231																		
	aattc	gggtg	gag	ctg	agcc	ggag	acagg	c	agtt	gtg	aaa	aact	tcag	ga	caaaa	atg	58		
																Met			
30																1			
	ttt	cat	tta	agg	act	tgt	gct	gct	aag	ttg	agg	cca	ttg	acg	gct	tcc	106		
	Phe	His	Leu	Arg	Thr	Cys	Ala	Ala	Lys	Leu	Arg	Pro	Leu	Thr	Ala	Ser			
				5					10					15					
	cag	act	gtt	aag	aca	ttt	tca	caa	aac	aga	cca	gca	gca	gct	agg	aca	154		
35	Gln	Thr	Val	Lys	Thr	Phe	Ser	Gln	Asn	Arg	Pro	Ala	Ala	Ala	Arg	Thr			
				20					25					30					
	ttt	caa	cag	att	cgt	gct	att	ctg	cac	ctg	ttg	ctg	ctg	agc	cct	ttc	202		
	Phe	Gln	Gln	Ile	Arg	Ala	Ile	Leu	His	Leu	Leu	Leu	Leu	Ser	Pro	Phe			
				35				40					45						
40	tca	gtg	gga	cta	gtt	cga	act	atg	tgg	agg	aga	tgt	act	gtg	ctt	ggc	250		
	Ser	Val	Gly	Leu	Val	Arg	Thr	Met	Trp	Arg	Arg	Cys	Thr	Val	Leu	Gly			
						55					60				65				
	tgg	aaa	acc	cca	aaa	gtg	tac	ata	aga	cag	ggt	ccc	act	gtt	gtc	cag	298		
	Trp	Lys	Thr	Pro	Lys	Val	Tyr	Ile	Arg	Gln	Gly	Pro	Thr	Val	Val	Gln			
45					70					75					80				
	gct	gga	gtg	cag	tgg	cgg	gat	ctc	ggc	ttg	ctg	caa	cct	ccg	act	cct	346		
	Ala	Gly	Val	Gln	Trp	Arg	Asp	Leu	Gly	Leu	Leu	Gln	Pro	Pro	Thr	Pro			
				85					90					95					
	ggg	ttc	aag	tgatt	cttat	gcct	cag	cct	ccc	gag	taac	tggg	act	aca			395		
50	Gly	Phe	Lys																
				100															
	ggt	gcac	gtc	acc	acg	cct	g	act	agttttt	gtattttt	tag	agat	ggg	atttt	actttt		455		
	gtt	ggc	cagg	ctg	gtct	tga	acc	cct	ggcc	tca	agt	gat	c	accc	acctt	ggc	ctcc	caa	515
	agt	gct	ggga	ttac	agg	tat	gat	ca	aacc	ac	gcct	ggcc	at	gtcat	gcctt	gtg	acaga	aat	575
55	tcctt	tattc	tg	ttt	tgag	c	caata	aatat	ttat	aggtt	tcg	aaaa	aaaa	aaaa	aaaa		634		
	<210> 232																		
	<211> 583																		
	<212> DNA																		
60	<213> Homo sapiens																		
	<220>																		
	<221> CDS																		

<222> 76..498

<400> 232

```

aatatagcca gccgcggctg cccttgcgct tcccagagctg gcgggggtccg tgggtgcggga      60
5  tcgagattgc gggct atg gcg ccg aag gtt ttt cgt cag tac tgg gat atc      111
    Met Ala Pro Lys Val Phe Arg Gln Tyr Trp Asp Ile
      1          5          10
ccc gat ggc acc gat tgc cac cgc aaa gcc tac agc acc acc agt att      159
Pro Asp Gly Thr Asp Cys His Arg Lys Ala Tyr Ser Thr Thr Ser Ile
10  15          20          25
gcc agc gtc gct ggc ctg acc gcc gct gcc tac aga gtc aca ctc aat      207
Ala Ser Val Ala Gly Leu Thr Ala Ala Tyr Arg Val Thr Leu Asn
    30          35          40
cct ccg ggc acc ttc ctt gaa gga gtg gct aag gtt gga caa tac acg      255
15  Pro Pro Gly Thr Phe Leu Glu Gly Val Ala Lys Val Gly Gln Tyr Thr
    45          50          55          60
ttc act gca gct gct gtc ggg gcc gtg ttt ggc ctc acc acc tgc atc      303
Phe Thr Ala Ala Val Gly Ala Val Phe Gly Leu Thr Thr Cys Ile
    65          70          75
agc gcc cat gtc cgc gag aag ccc gac gac ccc ctg aac tac ttc ctc      351
20  Ser Ala His Val Arg Glu Lys Pro Asp Asp Pro Leu Asn Tyr Phe Leu
    80          85          90
ggg ggc tgc gcc gga ggc ctg act ctg gga gca cgc acg cac aac tac      399
Gly Gly Cys Ala Gly Gly Leu Thr Leu Gly Ala Arg Thr His Asn Tyr
25  95          100          105
ggg att ggc gcc gcc gcc tgc gtg tac ttt ggc ata gcg gcc tcc ctg      447
Gly Ile Gly Ala Ala Ala Cys Val Tyr Phe Gly Ile Ala Ala Ser Leu
    110          115          120
gtc aag atg ggc cgg ctg gag gcc tgg gag gtg ttt gca aaa ccc aag      495
30  Val Lys Met Gly Arg Leu Glu Gly Trp Glu Val Phe Ala Lys Pro Lys
    125          130          135          140
gtg tgagccctgt gcctgccggg acctccagcc tgcagaatgc gtccagaaat      548
Val
aaattctgtg tctgtgtgaa aaaaaaaaaa aaaaa      583
35

```

<210> 233

<211> 753

<212> DNA

<213> Homo sapiens

40

<220>

<221> CDS

<222> 199..600

45 <400> 233

```

atttttccga tgccaggcac cctcaaggca cagaggctgg ggctcatgtt gggggcactt      60
ggcctctcca ggccctgaag gcttctctgg gctgatgcga gctggggaac gggagggacg      120
gacgtgggag cgagaacgtc acactggagg cagctggtgg cacgatgggg gacagagtga      180
aagagccttc gtgtcacc atg gcc aca cac ccc gat ggc ttc cgg ctt gag      231
50  Met Ala Thr His Pro Asp Gly Phe Arg Leu Glu
    1          5          10
gga ccc ctg gct gca gcc cac agc cct ggg cct tgc act gtg ctc tac      279
Gly Pro Leu Ala Ala Ala His Ser Pro Gly Pro Cys Thr Val Leu Tyr
    15          20          25
gaa ggc cct gtc cgt ggg ctc tgc ccy ttt gcc ccg cga aat tcc aac      327
55  Glu Gly Pro Val Arg Gly Leu Cys Pro Phe Ala Pro Arg Asn Ser Asn
    30          35          40
acc atg gcg gcg gct gcc ctg gct gcc ccc agc ctg ggc ttc gat ggg      375
Thr Met Ala Ala Ala Ala Leu Ala Ala Pro Ser Leu Gly Phe Asp Gly
60  45          50          55
gtg att ggg gtg ctc gtg gct gat acc agc ctc acg gac atg cac gtg      423
Val Ile Gly Val Leu Val Ala Asp Thr Ser Leu Thr Asp Met His Val
    60          65          70          75

```

gtg gat gta gag ctg agc gga ccc cgg ggc ccc act ggc cga agc ttt 471
 Val Asp Val Glu Leu Ser Gly Pro Arg Gly Pro Thr Gly Arg Ser Phe
 80 85 90
 gct gtg cac acc cgc aga gag aac cct gcc gag cca ggc gcg gtc acc 519
 5 Ala Val His Thr Arg Arg Glu Asn Pro Ala Glu Pro Gly Ala Val Thr
 95 100 105
 ggc tcc gcc acc gtc acg gcc ttc tgg cgg agc ctc ctg gcc tgc tgc 567
 Gly Ser Ala Thr Val Thr Ala Phe Trp Arg Ser Leu Leu Ala Cys Cys
 110 115 120
 10 cag ctc ccc tcc agg ccg ggg atc cat ctc tgc tgagaagcct cctccctccc 620
 Gln Leu Pro Ser Arg Pro Gly Ile His Leu Cys
 125 130
 gagacaagat catctgcctg gcctctcacc accaccatcc caccctgcc ctgccccact 680
 tccccagggt ctcccttctg actcagtaaa gatcaccgct gcctcccccc gcaaataaaa 740
 15 aaaaaaaaaa aaa 753

 <210> 234
 <211> 762
 <212> DNA
 20 <213> Homo sapiens

 <220>
 <221> CDS
 <222> 211..612
 25
 <400> 234
 atttccgatg ccaggcacc ccaaggcaca gaggctgggg ctcattgttg gggcacttgg 60
 cctctccagg cctcgaaggc ttctcctggg ctgatgagag ctggggaacg ggaggacgg 120
 acgtgggagc gagaacgtca cactggaggc agctgggtgg acgatggggg acagagtga 180
 30 aggtagcaag tcaagagcct tcgtgtcacc atg gcc aca cac ccc gat ggc ttc 234
 Met Ala Thr His Pro Asp Gly Phe
 1 5
 cgg ctt gag gga ccc ctg gct gca gcg cac agc cct ggg cct tgc act 282
 Arg Leu Glu Gly Pro Leu Ala Ala His Ser Pro Gly Pro Cys Thr
 10 15 20
 35 gtg ctc tac gaa ggc cct gtc cgt ggg ctc tgc ccc ttt gcc ccg cga 330
 Val Leu Tyr Glu Gly Pro Val Arg Gly Leu Cys Pro Phe Ala Pro Arg
 25 30 35 40
 40 aat tcc aac acc atg tcg gcg gct gcc ctg gct gcc ccc agc ctg ggc 378
 Asn Ser Asn Thr Met Ser Ala Ala Ala Leu Ala Ala Pro Ser Leu Gly
 45 50 55
 ttc gat ggg gtg att ggg gtg ctc gtg gct gat acc agc ctc acg gac 426
 Phe Asp Gly Val Ile Gly Val Leu Val Ala Asp Thr Ser Leu Thr Asp
 60 65 70
 45 atg cac gtg gtg gat gta gag ctg agc gga ccc cgg ggc ccc acg tgc 474
 Met His Val Val Asp Val Glu Leu Ser Gly Pro Arg Gly Pro Thr Cys
 75 80 85
 cga agc ttt gct gtg cac acc cgc aga gag aac cct gcc gag cca ggc 522
 Arg Ser Phe Ala Val His Thr Arg Arg Glu Asn Pro Ala Glu Pro Gly
 90 95 100
 50 gcg gtc acc ggc tcc gcc acc gtc acg gcc ttc tgg cgg agc ctc ctg 570
 Ala Val Thr Gly Ser Ala Thr Val Thr Ala Phe Trp Arg Ser Leu Leu
 105 110 115 120
 gcc tgc tgc cag ctc ccc tcc agg ccg ggg atc cat ctc tgc 612
 55 Ala Cys Cys Gln Leu Pro Ser Arg Pro Gly Ile His Leu Cys
 125 130
 tgagaagcct cctccctccc gagacaagat catctgcctg gcctctcacc accaccatcc 672
 caccctgcc ctgccccact tccccagggt ctcccttctg actcagtaaa gatcaccgct 732
 gcctcccccc gccaaaaaaaa aaaaaaaaaa 762
 60
 <210> 235
 <211> 537
 <212> DNA

<213> Homo sapiens

<220>

<221> CDS

5 <222> 5..259

<400> 235

```

aaaa atg cta aag gta gaa gca act ggt agt ccc gag gaa ggg tgg gcg      49
      Met Leu Lys Val Glu Ala Thr Gly Ser Pro Glu Glu Gly Trp Ala
10      1      5      10      15
      ggt gga gag ccc cgg act gga gct cct gcg aac tcc cct tcc tgc cct      97
      Gly Gly Glu Pro Arg Thr Gly Ala Pro Ala Asn Ser Pro Ser Cys Pro
      20      25      30
      cag gag atg cca ctg cag gac cca agg agc agg gag gag gcg gcc agg      145
15      Gln Glu Met Pro Leu Gln Asp Pro Arg Ser Arg Glu Glu Ala Ala Arg
      35      40      45
      acc cag cag cta ttg ctg gcc act ctg cag gag gca gcg acc acg cag      193
      Thr Gln Gln Leu Leu Leu Ala Thr Leu Gln Glu Ala Ala Thr Thr Gln
      50      55      60
      gag aac gtg gcc tgg agg aag aac tgg atg gtt ggc ggc gaa ggc ggc      241
20      Glu Asn Val Ala Trp Arg Lys Asn Trp Met Val Gly Gly Glu Gly Gly
      65      70      75
      gcc agc ggg agg tca ccg tgagaccgga cttgcctccg tgggcgcccgg      289
      Ala Ser Gly Arg Ser Pro
25      80      85
      accttggtt gggcgcagga atccgaggca gcctttctcc ttcgtgggcc cagcggagag      349
      tccggaccga gataccatgc caggactctc cggggctctg tgagctgccg tcgggtgagc      409
      acgtttcccc caaaccttg actgactgct ttaagggtccg caaggcgggc cagggccgag      469
      acgcgagtcg gatgtggtga actgaaagaa ccaataaaat catgttcctc caaaaaaaaa      529
30      aaaaaaaaaa      537

```

<210> 236

<211> 994

<212> DNA

35 <213> Homo sapiens

<220>

<221> CDS

<222> 23..370

40

<400> 236

```

gattgctggt tgctgtaaag tg atg ggg agg ccc tgg atg gtg atg ata ttg      52
      Met Gly Arg Pro Trp Met Val Met Ile Leu
      1      5      10
45      gag tca aaa tct gaa gaa aag atg tgg tat ggt gta ttc ctg tgg gca      100
      Glu Ser Lys Ser Glu Glu Lys Met Trp Tyr Gly Val Phe Leu Trp Ala
      15      20      25
      ctg gtg tct tct ctc ttc ttt cat gtc cct gct gga tta ctg gcc ctc      148
50      Leu Val Ser Ser Leu Phe Phe His Val Pro Ala Gly Leu Leu Ala Leu
      30      35      40
      ttc acc ctc aga cat cac aaa tat ggt agg ttc atg tct gta agc atc      196
      Phe Thr Leu Arg His His Lys Tyr Gly Arg Phe Met Ser Val Ser Ile
      45      50      55
      ctg ttg atg ggc atc gtg gga cca att act gct gga atc ttg aca agt      244
55      Leu Leu Met Gly Ile Val Gly Pro Ile Thr Ala Gly Ile Leu Thr Ser
      60      65      70
      gca gct att gct gga gtt tac cga gca gca ggg aag gaa atg ata cca      292
      Ala Ala Ile Ala Gly Val Tyr Arg Ala Ala Gly Lys Glu Met Ile Pro
      75      80      85      90
60      ttt gaa gcc ctc aca ctg ggc act gga cag aca ttt tgc gtc ttg gtg      340
      Phe Glu Ala Leu Thr Leu Gly Thr Gly Gln Thr Phe Cys Val Leu Val
      95      100      105
      gtc tcc ttt tta cgg att tta gct act cta tagcatatcat ccttatgctg      390

```

Val Ser Phe Leu Arg Ile Leu Ala Thr Leu									
			110				115		
	agatggttgaa	cttaaacttt	atggaatcct	ccaaaagaat	acattatgga	gtgtagtggt	450		
	ttcttagttc	ttcaaaggga	agcaacttgg	atgaacagga	acatgaagga	caacacatct	510		
5	cagccttttc	ttcattttga	agctcctaga	attgaagact	tatgtggact	cctattgttc	570		
	tcaacccaaaa	caagtctttt	ggctttcttt	tttgtagata	tttaatttaa	gcagttttca	630		
	tgtgtacctt	tacccaagcc	aagtcaacag	tgtctctggg	gtggcaccct	ttgcactgaa	690		
	atttacagta	ttctgtgaga	tgctgcataat	tttgaagaaa	ccgtggaaga	tactgggtta	750		
	tttcaaataa	gcagagtatg	ttgtattaaa	atcttatcta	atcttgatta	aaatttgga	810		
10	aactcttttc	tttgctacat	cttagtgaca	ataaatgcca	aatagggttt	ggttgagtat	870		
	agttttgaaa	acaaatttgg	tgaaataaag	caggaaaaaa	aattttaagta	taactcaagt	930		
	agtggtcttg	gttccactgt	ttataaataa	aaagtagata	acaatggaaa	aaaaaaaaaa	990		
	aaaa						994		
15	<210> 237								
	<211> 662								
	<212> DNA								
	<213> Homo sapiens								
20	<220>								
	<221> CDS								
	<222> 41..352								
	<400> 237								
25	tagctaaaaa	ttgagggttc	taaataactaa	ggaagaaggg	atg aat aga tat tgt	55			
					Met Asn Arg Tyr Cys				
					1 5				
	ggc aag ata ttt gtc tct gtc atg gtt aaa ttg caa aaa aat aaa ctt	103							
	Gly Lys Ile Phe Val Ser Val Met Val Lys Leu Gln Lys Asn Lys Leu								
30		10		15		20			
	acc tcc ttc ccc agg cag cca ttg tta aca ttt ttt gaa tat cta gaa	151							
	Thr Ser Phe Pro Arg Gln Pro Leu Leu Thr Phe Phe Glu Tyr Leu Glu								
		25		30		35			
	aaa gtc ctt tgt tca gga tta ttt tcc cac tct gcc aag agt cac cat	199							
35	Lys Val Leu Cys Ser Gly Leu Phe Ser His Ser Ala Lys Ser His His								
		40		45		50			
	gac ctg ctc aca cgc cac cct tat gaa act gcc gcg cca ctt ctc agc	247							
	Asp Leu Leu Thr Arg His Pro Tyr Glu Thr Ala Ala Pro Leu Leu Ser								
		55		60		65			
40	tcc cat ttg att ctc aca gaa gct cta cga aat ggg ttg ggc aaa tgt	295							
	Ser His Leu Ile Leu Thr Glu Ala Leu Arg Asn Gly Leu Gly Lys Cys								
		70		75		80		85	
	cat gat cct cat ttc aca ggg gaa gaa act gag gcc cag agg ggg aaa	343							
	His Asp Pro His Phe Thr Gly Glu Glu Thr Glu Ala Gln Arg Gly Lys								
45		90		95		100			
	ctg act acc taaaattgcc atgtaggccg gcgcggtggc tcacgcctgt	392							
	Leu Thr Thr								
	aatcccagca ctgtgggagg ccaaggcggg tggatcgca ggtcaggaga tcgagaccat	452							
	cctgggtggc acttgaagcc ccgtctctac tagggataca aataattggc cgggtgtggg	512							
50	ggcgggcgcc tgtgkwccca gctgttcggc aggctgagga gggcgaaatgg tgtgagcctg	572							
	ggagggcgag cttgcggtgg gccgggattg cgccactgca ctccagcctg ggcgacagag	632							
	ccagattccg tccaaaaaaa aaaaaaaaaa	662							
	<210> 238								
55	<211> 1829								
	<212> DNA								
	<213> Homo sapiens								
	<220>								
60	<221> CDS								
	<222> 3..1319								
	<400> 238								

	at	cta	ggt	gac	cat	gga	tgg	gag	ctg	agc	ttg	gag	gag	gac	gca	cag	47
	Leu	Gly	Asp	His	Gly	Trp	Glu	Leu	Ser	Leu	Glu	Glu	Asp	Ala	Gln		
	1				5					10					15		
	ctg	tgg	ggt	ggg	gtg	gtg	aag	agt	tgt	ttt	gag	gga	aaa	ggc	cca	caa	95
5	Leu	Trp	Gly	Gly	Val	Val	Lys	Ser	Cys	Phe	Glu	Gly	Lys	Gly	Pro	Gln	
					20					25					30		
	aga	gaa	gcc	caa	cca	gcc	agc	ccc	cag	gcc	gcc	ccg	cca	gga	ccc	acc	143
	Arg	Glu	Ala	Gln	Pro	Ala	Ser	Pro	Gln	Ala	Ala	Pro	Pro	Gly	Pro	Thr	
				35					40					45			
10	aat	gag	gca	cag	atg	gca	gcc	gct	gcc	gcc	cta	gcc	cgg	ctg	gag	cag	191
	Asn	Glu	Ala	Gln	Met	Ala	Ala	Ala	Ala	Ala	Leu	Ala	Arg	Leu	Glu	Gln	
				50				55					60				
	aag	cag	tcc	cgg	gcc	tgg	ggc	ccc	aca	tcg	cag	gac	acc	atc	cga	aac	239
	Lys	Gln	Ser	Arg	Ala	Trp	Gly	Pro	Thr	Ser	Gln	Asp	Thr	Ile	Arg	Asn	
15		65					70					75					
	cag	gtg	aga	aag	gaa	ctt	caa	gcc	gaa	gcc	acc	gtc	agc	ggg	agc	ccc	287
	Gln	Val	Arg	Lys	Glu	Leu	Gln	Ala	Glu	Ala	Thr	Val	Ser	Gly	Ser	Pro	
					80		85				90				95		
	gag	gcc	cca	ggg	acc	aac	gtg	gta	tct	gag	ccc	aga	gag	gaa	ggc	tct	335
20	Glu	Ala	Pro	Gly	Thr	Asn	Val	Val	Ser	Glu	Pro	Arg	Glu	Glu	Gly	Ser	
					100					105					110		
	gcc	cac	ctg	gct	gtg	cct	ggc	gtg	tac	ttc	acc	tgt	ccg	ctc	act	ggg	383
	Ala	His	Leu	Ala	Val	Pro	Gly	Val	Tyr	Phe	Thr	Cys	Pro	Leu	Thr	Gly	
				115					120					125			
25	gcc	acc	ctg	agg	aag	gac	cag	cgg	gac	gcc	tgc	atc	aag	gag	gcc	att	431
	Ala	Thr	Leu	Arg	Lys	Asp	Gln	Arg	Asp	Ala	Cys	Ile	Lys	Glu	Ala	Ile	
				130				135					140				
	ctc	ttg	cac	ttc	tcc	acc	gac	cca	gtg	gcc	gcc	tcc	atc	atg	aag	atc	479
	Leu	Leu	His	Phe	Ser	Thr	Asp	Pro	Val	Ala	Ala	Ser	Ile	Met	Lys	Ile	
30		145					150					155					
	tac	acg	ttc	aac	aaa	gac	cag	gac	cgg	gtg	aag	ctg	ggt	gtg	gac	acc	527
	Tyr	Thr	Phe	Asn	Lys	Asp	Gln	Asp	Arg	Val	Lys	Leu	Gly	Val	Asp	Thr	
		160				165				170					175		
	att	gcc	aag	tac	ctg	gac	aac	atc	cac	ctg	cac	ccc	gag	gag	gag	aag	575
35	Ile	Ala	Lys	Tyr	Leu	Asp	Asn	Ile	His	Leu	His	Pro	Glu	Glu	Glu	Lys	
					180					185					190		
	tac	cgg	aag	atc	aag	ctg	cag	aac	aag	gtg	ttt	cag	gag	cgc	att	aac	623
	Tyr	Arg	Lys	Ile	Lys	Leu	Gln	Asn	Lys	Val	Phe	Gln	Glu	Arg	Ile	Asn	
				195					200					205			
40	tgc	ctg	gaa	ggg	acc	cac	gag	ttt	ttt	gag	gcc	att	ggg	ttc	cag	aag	671
	Cys	Leu	Glu	Gly	Thr	His	Glu	Phe	Phe	Glu	Ala	Ile	Gly	Phe	Gln	Lys	
				210				215					220				
	gtg	ttg	ctt	ccc	gcc	cag	gat	cag	gag	gac	ccc	gag	gag	ttc	tac	gtg	719
	Val	Leu	Leu	Pro	Ala	Gln	Asp	Gln	Glu	Asp	Pro	Glu	Glu	Phe	Tyr	Val	
45		225					230					235					
	ctg	agc	gag	acc	acc	ttg	gcc	cag	ccc	cag	agc	ctg	gag	agg	cac	aag	767
	Leu	Ser	Glu	Thr	Thr	Leu	Ala	Gln	Pro	Gln	Ser	Leu	Glu	Arg	His	Lys	
						240		245			250				255		
	gaa	cag	ctg	ctg	gct	gcg	gag	ccc	gtg	cgc	gcc	aag	ctg	gac	agg	cag	815
50	Glu	Gln	Leu	Leu	Ala	Ala	Glu	Pro	Val	Arg	Ala	Lys	Leu	Asp	Arg	Gln	
					260					265					270		
	cgc	cgc	gtc	ttc	cag	ccc	tcg	ccc	ctg	gcc	tcg	cag	ttc	gaa	ctg	cct	863
	Arg	Arg	Val	Phe	Gln	Pro	Ser	Pro	Leu	Ala	Ser	Gln	Phe	Glu	Leu	Pro	
				275					280					285			
55	ggg	gac	ttc	ttc	aac	ctc	aca	gca	gag	gag	atc	aag	cgg	gag	cag	agg	911
	Gly	Asp	Phe	Phe	Asn	Leu	Thr	Ala	Glu	Glu	Ile	Lys	Arg	Glu	Gln	Arg	
				290				295					300				
	ctc	agg	tcc	gag	gcg	gtg	gag	cgg	ctg	agc	gtg	ctg	cgg	acc	aag	gcc	959
	Leu	Arg	Ser	Glu	Ala	Val	Glu	Arg	Leu	Ser	Val	Leu	Arg	Thr	Lys	Ala	
60		305					310					315					
	atg	cgg	gag	aag	gag	gag	cag	cgg	ggg	ctg	cgc	aag	tac	aac	tac	acg	1007
	Met	Arg	Glu	Lys	Glu	Glu	Gln	Arg	Gly	Leu	Arg	Lys	Tyr	Asn	Tyr	Thr	
		320				325					330				335		

	ctg ctg cgc gtg cgc ctc ccc gat ggc tgc ctc ctg cag ggc act ttc	1055
	Leu Leu Arg Val Arg Leu Pro Asp Gly Cys Leu Leu Gln Gly Thr Phe	
	340 345 350	
5	tac gct cgg gag cgg ctg ggg gcg gtg tac ggg ttc gtc cgg gag gcc	1103
	Tyr Ala Arg Glu Arg Leu Gly Ala Val Tyr Gly Phe Val Arg Glu Ala	
	355 360 365	
	ctg cag agc gac tgg ctg cct ttt gag ctg ctg gcc tgc gga ggg cag	1151
	Leu Gln Ser Asp Trp Leu Pro Phe Glu Leu Leu Ala Ser Gly Gly Gln	
	370 375 380	
10	aag ctg tcc gag gac gag aac ctg gcc ttg aac gag tgc ggg ctg gtg	1199
	Lys Leu Ser Glu Asp Glu Asn Leu Ala Leu Asn Glu Cys Gly Leu Val	
	385 390 395	
	ccc tct gcc ctc ctg acc ttc tgc tgg gac atg gct gtg ctg gag gac	1247
	Pro Ser Ala Leu Leu Thr Phe Ser Trp Asp Met Ala Val Leu Glu Asp	
15	400 405 410 415	
	atc aag gcc gcg ggg gcc gag ccg gac tcc atc ctg aaa ccc gag ctc	1295
	Ile Lys Ala Ala Gly Ala Glu Pro Asp Ser Ile Leu Lys Pro Glu Leu	
	420 425 430	
20	ctg tca gcc atc gag aag ctc ttg tgaaataaaa gcagggttgg cctcagccct	1349
	Leu Ser Ala Ile Glu Lys Leu Leu	
	435	
	gtgggtctgt ctcattgctct cctgttctt ctccccgccca cccaggggcc tccaagccac	1409
	ctctggaaat acttggctct gccccatggg cacgggaggg gcgccagccg tggagctgtg	1469
25	gaattggggc ccgtggcaga gccccatcc cttgggggct gtggggatgc gccaagccc	1529
	ccgagggaga ggcctgggga caccaacaaa tctaagccct ccctagctct tggtaactgt	1589
	gtcatgaagc tgccggacag acacacgtgg catctccctg ggcaggagag caggcctgca	1649
	gcatgggtcc cgttcccgtg tgccgtgggt ggcagtggct gcacctggca ctagggctgc	1709
	tctgtggatg tgggtgacaa cggcaggagg ggacgctggc cttcctgcac atagacctgc	1769
	agttagtaaa tcataagccc aaataaacag gttgtttgaa tataaaaaaa aaaaaaaaaa	1829
30	<210> 239	
	<211> 1083	
	<212> DNA	
	<213> Homo sapiens	
35	<220>	
	<221> CDS	
	<222> 421..768	
40	<400> 239	
	aaggatgtgc tctttcccaa ggagagggag ctctgttgcc tcttcccccac agaatcactc	60
	tgtgcaaacc tcttcccctc cttggcccca gtctcccccac ttctaaaatc ggatactgga	120
	taaaatgcc aaggaagaacc tagggatgca ccaggaacca cgcgcctgaa tgccacaggt	180
	ttgatttgat tcatgaccct catctggaca caagctctaa aatacttgag ccttggcaga	240
45	aatggctgat agagtccaca gaacacgctg tcctcatctc agagaggaga actctgaacc	300
	cagaggggaa ggatttacct gcagttgtat ggcaagccag aggtaggcgc tgcactggaa	360
	cgcagcctaa ccagcctaaa gaaacatgg gaggagaggc tcttaccctc tcctttgcag	420
	atg tgg gcc cgg ctg cct cac act cca gag cag atg ggc cac agg ctt	468
	Met Trp Ala Arg Leu Pro His Thr Pro Glu Gln Met Gly His Arg Leu	
50	1 5 10 15	
	ata ggt ccc aag gaa gct tca ctt cat gtg gta ccc agc tgg cca gcc	516
	Ile Gly Pro Lys Glu Ala Ser Leu His Val Val Pro Ser Trp Pro Ala	
	20 25 30	
	agg aag atg gag ggg ctt ctg gct ggc ctc tct tcc tct cct aga aag	564
55	Arg Lys Met Glu Gly Leu Leu Ala Gly Leu Ser Ser Ser Pro Arg Lys	
	35 40 45	
	tca tgc tgg ccc ttt tgg gtc cat ggg cca aag gtt cat gaa ggt ggc	612
	Ser Cys Trp Pro Phe Trp Val His Gly Pro Lys Val His Glu Gly Gly	
	50 55 60	
60	tct gcc tgt gag aca tca agc tcc tgg gtt gaa gga ctt gga tta aga	660
	Ser Ala Cys Glu Thr Ser Ser Ser Trp Val Glu Gly Leu Gly Leu Arg	
	65 70 75 80	
	aga gtg aca tca gtg cac agt tta tgc caa ggg ctt ggg gcc tca gtc	708

	Arg	Val	Thr	Ser	Val	His	Ser	Leu	Cys	Gln	Gly	Leu	Gly	Ala	Ser	Val	
					85					90					95		
	cag	ctt	ctt	cct	gga	cca	cca	cca	aca	aca	acc	agt	gat	aaa	aat	aat	756
	Gln	Leu	Leu	Pro	Gly	Pro	Pro	Pro	Thr	Thr	Thr	Ser	Asp	Lys	Asn	Asn	
5				100					105					110			
	tat	act	agt	ggc	tgacatttat	ggattcttcc	tacacactag	gctataaccac									808
	Tyr	Thr	Ser	Gly													
				115													
	agcgagtgcc	tcgaaaggaa	atatagtata	gcactgtgcc	gtccaacatg	gcgggccacta											868
10	gccacatgca	ctactgagca	cttgaaatgt	ggctagccca	cattgagatg	tgctgtaaat											928
	aaagaataga	caccagattt	ccaagactta	gtaccaaaaa	aagaatgtaa	aattttctcat											988
	taacaatttt	ttttcttaca	tttattacat	gttaacatga	cgctatttgg	agttttaaata											1048
	aatgcattat	taaaattcaa	aaaaaaaaaa	aaaaa													1083
15	<210>	240															
	<211>	1831															
	<212>	DNA															
	<213>	Homo sapiens															
20	<220>																
	<221>	CDS															
	<222>	78..590															
	<400>	240															
25	aaggacttaa	gcgccccgga	gccgggaggc	gaacttggga	cccgcctggcc	tcgctcggtg											60
	cgcgctctcc	tccccgc	atg	cag	ccc	gcc	gag	cgc	tcg	cgg	gtc	ccc	agg				110
			Met	Gln	Pro	Ala	Glu	Arg	Ser	Arg	Val	Pro	Arg				
			1				5					10					
	atc	gac	ccg	tac	gga	ttc	gag	cgg	cct	gag	gac	ttc	gac	gac	gcc	gcc	158
30	Ile	Asp	Pro	Tyr	Gly	Phe	Glu	Arg	Pro	Glu	Asp	Phe	Asp	Asp	Ala	Ala	
			15				20					25					
	tac	gag	aag	ttt	ttc	tcc	agc	tac	ctg	gtc	acg	ctc	acc	cgc	agg	gcg	206
	Tyr	Glu	Lys	Phe	Phe	Ser	Ser	Tyr	Leu	Val	Thr	Leu	Thr	Arg	Arg	Ala	
			30				35					40					
35	atc	aaa	tgg	tcc	cgg	ctg	ctg	cag	ggc	ggg	ggc	gtc	ccc	agg	agc	cgg	254
	Ile	Lys	Trp	Ser	Arg	Leu	Gln	Gly	Gly	Gly	Val	Pro	Arg	Ser	Arg		
		45				50					55						
	aca	gtg	aag	cgc	tat	gtc	cgg	aaa	ggg	gtc	ccg	ctg	gag	cac	cgt	gcc	302
	Thr	Val	Lys	Arg	Tyr	Val	Arg	Lys	Gly	Val	Pro	Leu	Glu	His	Arg	Ala	
40	60					65				70					75		
	cgc	gtc	tgg	atg	gtg	ctg	agt	ggg	gcc	cag	gcg	cag	atg	gac	cag	aat	350
	Arg	Val	Trp	Met	Val	Leu	Ser	Gly	Ala	Gln	Ala	Gln	Met	Asp	Gln	Asn	
				80						85					90		
	ccc	ggc	tac	tac	cac	cag	ctt	ctc	cag	gga	gag	aga	aac	ccc	agg	ctg	398
45	Pro	Gly	Tyr	Tyr	His	Gln	Leu	Leu	Gln	Gly	Glu	Arg	Asn	Pro	Arg	Leu	
			95				100						105				
	gag	gac	gcc	atc	agg	aca	gac	ctg	aac	cgg	acc	ttc	ccc	gac	aac	gtg	446
	Glu	Asp	Ala	Ile	Arg	Thr	Asp	Leu	Asn	Arg	Thr	Phe	Pro	Asp	Asn	Val	
			110				115					120					
50	aag	ttc	cgg	aag	acc	acg	gac	ccc	tgc	tta	cag	agg	acc	ctg	tac	aat	494
	Lys	Phe	Arg	Lys	Thr	Thr	Asp	Pro	Cys	Leu	Gln	Arg	Thr	Leu	Tyr	Asn	
			125				130					135					
	gtg	ctg	ctg	gca	tat	ggg	cac	cat	aac	cag	gga	gtg	ggc	tac	tgc	cag	542
	Val	Leu	Leu	Ala	Tyr	Gly	His	His	Asn	Gln	Gly	Val	Gly	Tyr	Cys	Gln	
55	140					145					150				155		
	gga	atg	aat	ttt	ata	gca	gga	tat	ctg	att	ctt	ata	aca	aat	aat	gaa	590
	Gly	Met	Asn	Phe	Ile	Ala	Gly	Tyr	Leu	Ile	Leu	Ile	Thr	Asn	Asn	Glu	
				160						165					170		
	taagaatctt	tttggctgtt	agatgctctt	gttgggaagaa	tactaccaga	ttactacagc											650
60	ccggccatgc	tgggcctgaa	gaccgaccag	gaggtcctcg	gggagctggt	gcgggcgaag											710
	ctgccggctg	tgggggccct	gatggagcgt	ctcgggtgtgc	tgtggacgct	gctgggtgtcc											770
	cgctggttca	tctgcctgtt	tgtggacatc	ttgcccgtgg	agacagtgtc	tcggatctgg											830
	gactgtttgt	ttaacgaagg	ctcgaagatt	atcttccggt	tggccctgac	cttaattaag											890

	cagcaccagg	agttgatttt	ggaagccacc	agcgttccag	acattttgcga	taagtttaag	950
	cagataacca	aagggagttt	cgtgatggag	tgtcacacgt	ttatgcaggt	gtgtggggct	1010
	gcacgtggct	cagtcctcctc	ccagggggcc	ccgcctcacc	tgacagcccg	gggctgctct	1070
	gaccaccccg	aggatgcaca	ggatggggcac	cagtgggcat	agggcacagg	atgagcctcc	1130
5	agctctgtcc	tgcattctgcc	ccctgcgcct	ggcctccgag	ggctttcctg	tctatggcgg	1190
	ccctgtcttc	ttggccctgg	caactgcggac	gctgtctcctg	gtcctaattg	ctgtactcat	1250
	ctgctgtgtg	tgggtgccaga	agtgtggctt	cccagaggcc	ggcctcccca	ctgggtcctg	1310
	gacctggcgc	aggccgtata	gactcagggtc	ctgatgaggg	cgttgtggga	gctgtacctg	1370
	acaggccttc	tgaggaagcc	aagacgccag	gagaggctca	ggcctgggag	tcagtagttt	1430
10	cctaagaggg	agtggaggct	cggggccact	ctgggtgcag	catggcaaac	gtgggcggta	1490
	tttcagcagc	tgggccttca	tcaaagagaa	gacctgttg	gccgggcgcg	gtggctcacg	1550
	cctgcagtc	cagcactttg	ggaggccaag	gcgtgtggat	cacctgaggt	caggagttca	1610
	agaccagcct	ggccaacacg	gtgaaacccc	gtctctacta	aaaaatacaa	aaattagcca	1670
	gggtgtgggtg	ctcacgctta	tgtagtccca	gttactcggg	aggctgaggg	acgagaatca	1730
15	cttgaacctg	ggaggcggag	gttgacagtg	gccgagatcg	cgccactgca	ctccagcctg	1790
	ggcaacagag	tgagactctg	tctcaaaaaa	aaaaaaaaaa	a		1831
	<210> 241						
	<211> 1830						
20	<212> DNA						
	<213> Homo sapiens						
	<220>						
	<221> CDS						
25	<222> 78..608						
	<400> 241						
	aaggacttaa	gcgccccgga	gccgggaggg	gaacttggga	cccgtctggc	tcgctcggtg	60
	cgcgcctccc	tccccgc	atg cag ccc gcc	gag cgc tcg cgg gtc	ccc agg		110
30	Met Gln Pro Ala Glu Arg Ser Arg Val Pro Arg						
	1 5 10						
	atc gac ccg tac gga ttc gag cgg cct gag gac ttc gac gac gcc gcc	158					
	Ile Asp Pro Tyr Gly Phe Glu Arg Pro Glu Asp Phe Asp Asp Ala Ala						
	15 20 25						
35	tac gag aag ttt ttc tcc agc tac ctg gtc acg ctc acc cgc agg gcg	206					
	Tyr Glu Lys Phe Phe Ser Ser Tyr Leu Val Thr Leu Thr Arg Arg Ala						
	30 35 40						
	atc aaa tgg tcc cgg ctg ctg cag ggc ggg ggc gtc ccc agg agc cgg	254					
	Ile Lys Trp Ser Arg Leu Leu Gln Gly Gly Gly Val Pro Arg Ser Arg						
40	45 50 55						
	aca gtg aag cgc tat gtc cgg aaa ggg gtc ccg ctg gag cac cgt gcc	302					
	Thr Val Lys Arg Tyr Val Arg Lys Gly Val Pro Leu Glu His Arg Ala						
	60 65 70 75						
	cgc gtc tgg atg gtg ctg agt ggg gcc cag gcg cag atg gac cag aat	350					
45	Arg Val Trp Met Val Leu Ser Gly Ala Gln Ala Gln Met Asp Gln Asn						
	80 85 90						
	ccc ggc tac tac cac cag ctt ctc cag gga gag aga aac ccc agg ctg	398					
	Pro Gly Tyr Tyr His Gln Leu Leu Gln Gly Glu Arg Asn Pro Arg Leu						
	95 100 105						
50	gag gac gcc atc agg aca gac ctg aac cgg acc ttc ccc gac aac gtg	446					
	Glu Asp Ala Ile Arg Thr Asp Leu Asn Arg Thr Phe Pro Asp Asn Val						
	110 115 120						
	aag ttc cgg aag acc acg gac ccc tgc tta cag agg acc ctg tac aat	494					
	Lys Phe Arg Lys Thr Thr Asp Pro Cys Leu Gln Arg Thr Leu Tyr Asn						
55	125 130 135						
	gtg ctg ctg gca tat ggg cac cat aac cag gga gtg ggc tac tgc cag	542					
	Val Leu Leu Ala Tyr Gly His His Asn Gln Gly Val Gly Tyr Cys Gln						
	140 145 150 155						
	gga atg aat ttt ata gca gga tat ctg att ctt ata aca aat aat gat	590					
60	Gly Met Asn Phe Ile Ala Gly Tyr Leu Ile Leu Ile Thr Asn Asn Asp						
	160 165 170						
	aag aat ctt ttt ggc tgt tagatgctct tgttggaaga atactaccag	638					
	Lys Asn Leu Phe Gly Cys						

175
 attactacag cccggccatg ctgggcctga agaccgacca ggaggtcctc ggggagctgg 698
 tgccggcgaa gctgccggct gtggggggccc tgatggagcg tctcggtgtg ctgtggacgc 758
 tgctgggtgtc ccgctgggttc atctgcctgt ttgtggacat cttgcccgtg gagacagtgc 818
 5 ttcggatctg ggactgtttg tttaacgaag gctcgaagat tatcttccgg ttggccctga 878
 ccttaattaa gcagaccag gagttgattt tggaagccac cagcgttcca gacatttgcg 938
 ataagtttaa gcagataacc aaaggagggt tcgtgatgga gtgtcacacg tttatgcagg 998
 tgtgtggggc tgcacgtggc tcagtcccct cccagggggc cccgcctcac ctgcagcccg 1058
 ggggctgctc tgaccacccg gaggatgcac aggatgggca ccagtgggca tagggcacag 1118
 10 gatgagcctc cagctctgtc ctgcatctgc cccctgcgcc tggcctccga gggctttcct 1178
 gtctatggcg gccctgtctt cttggccctg gactgcgga cgctgctcct ggtcctaattg 1238
 gctgtactca tctgtgtgt gtggtgccag aagtgtggct tcccgaggcc cggcctcccc 1298
 actgggtcct ggacctggcg caggccgtat agactcaggt cctgatgagg gcgttgtggg 1358
 agctgtacct gacaggcctt ctgaggaagc caagacgcca ggagaggctc aggcctggga 1418
 15 gtcagtagtt tcctaagagg gagtggaggc tcggggccac tctgggtgca gcatggcaaa 1478
 cgtgggcggg atttcagcag ctgggccttc atcaaagaga agaccatgtt ggccgggcgc 1538
 ggtggtcac gcctgcagtc ccagcacttt gggaggccaa ggcgtgtgga tcacctgagg 1598
 tcaggagttc aagaccagcc tggccaacac ggtgaaaccc cgtctctact aaaaaatata 1658
 aaaattagcc aggtgtggtg gctcacgctt atgtagtccc agttactcgg gaggtgagg 1718
 20 cacgagaatc acttgaacct gggaggcgga ggttgcagtg agccgagatc gcgccactgc 1778
 actccagcct gggcaacaga gtgagactct gtctcaaaaa aaaaaaaaaa aa 1830

<210> 242
 <211> 508
 25 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 30 <222> -27...-1

<400> 242
 Met Asp Pro Lys Leu Gly Arg Met Ala Ala Ser Leu Leu Ala Val Leu
 -25 -20 -15
 35 Leu Leu Leu Leu Leu Glu Arg Gly Met Phe Ser Ser Pro Ser Pro Pro
 -10 -5 1 5
 Pro Ala Leu Leu Glu Lys Val Phe Gln Tyr Ile Asp Leu His Gln Asp
 10 15 20
 40 Glu Phe Val Gln Thr Leu Lys Glu Trp Val Ala Ile Glu Ser Asp Ser
 25 30 35
 Val Gln Pro Val Pro Arg Phe Arg Gln Glu Leu Phe Arg Met Met Ala
 40 45 50
 Val Ala Ala Asp Thr Leu Gln Arg Leu Gly Ala Arg Val Ala Ser Val
 55 60 65
 45 Asp Met Gly Pro Gln Gln Leu Pro Asp Gly Gln Ser Leu Pro Ile Pro
 70 75 80 85
 Pro Val Ile Leu Ala Glu Leu Gly Ser Asp Pro Thr Lys Gly Thr Val
 90 95 100
 50 Cys Phe Tyr Gly His Leu Asp Val Gln Pro Ala Asp Arg Gly Asp Gly
 105 110 115
 Trp Leu Thr Asp Pro Tyr Val Leu Thr Glu Val Asp Gly Lys Leu Tyr
 120 125 130
 Gly Arg Gly Ala Thr Asp Asn Lys Gly Pro Val Leu Ala Trp Ile Asn
 135 140 145
 55 Ala Val Ser Ala Phe Arg Ala Leu Glu Gln Asp Leu Pro Val Asn Ile
 150 155 160 165
 Lys Phe Ile Ile Glu Gly Met Glu Glu Ala Gly Ser Val Ala Leu Glu
 170 175 180
 60 Glu Leu Val Glu Lys Glu Lys Asp Arg Phe Phe Ser Gly Val Asp Tyr
 185 190 195
 Ile Val Ile Ser Asp Asn Leu Trp Ile Ser Gln Arg Lys Pro Ala Ile
 200 205 210
 Thr Tyr Gly Thr Arg Gly Asn Ser Tyr Phe Met Val Glu Val Lys Cys

215 220 225
 Arg Asp Gln Asp Phe His Ser Gly Thr Phe Gly Gly Ile Leu His Glu
 230 235 240 245
 Pro Met Ala Asp Leu Val Ala Leu Leu Gly Ser Leu Val Asp Ser Ser
 5 Gly His Ile Leu Val Pro Gly Ile Tyr Asp Glu Val Val Pro Leu Thr
 250 255 260
 265 270 275
 Glu Glu Glu Ile Asn Thr Tyr Lys Ala Ile His Leu Asp Leu Glu Glu
 280 285 290
 10 Tyr Arg Asn Ser Ser Arg Val Glu Lys Phe Leu Phe Asp Thr Lys Glu
 295 300 305
 Glu Ile Leu Met His Leu Trp Arg Tyr Pro Ser Leu Ser Ile His Gly
 310 315 320 325
 Ile Glu Gly Ala Phe Asp Glu Pro Gly Thr Lys Thr Val Ile Pro Gly
 15 330 335 340
 Arg Val Ile Gly Lys Phe Ser Ile Arg Leu Val Pro His Met Asn Val
 345 350 355
 Ser Ala Val Glu Lys Gln Val Thr Arg His Leu Glu Asp Val Phe Ser
 360 365 370
 20 Lys Arg Asn Ser Ser Asn Lys Met Val Val Ser Met Thr Leu Gly Leu
 375 380 385
 His Pro Trp Ile Ala Asn Ile Asp Asp Thr Gln Tyr Leu Ala Ala Lys
 390 395 400 405
 Arg Ala Ile Arg Thr Val Phe Gly Thr Glu Pro Asp Met Ile Arg Asp
 25 410 415 420
 Gly Ser Thr Ile Pro Ile Ala Lys Met Phe Gln Glu Ile Val His Lys
 425 430 435
 Ser Val Val Leu Ile Pro Leu Gly Ala Val Asp Asp Gly Glu His Ser
 440 445 450
 30 Gln Asn Glu Lys Ile Asn Arg Trp Asn Tyr Ile Glu Gly Thr Lys Leu
 455 460 465
 Phe Ala Ala Phe Phe Leu Glu Met Ala Gln Leu His
 470 475 480

 35 <210> 243
 <211> 331
 <212> PRT
 <213> Homo sapiens

 40 <220>
 <221> SIGNAL
 <222> -31...-1

 <400> 243
 45 Met Trp Leu Trp Glu Asp Gln Gly Gly Leu Leu Gly Pro Phe Ser Phe
 -30 -25 -20
 Leu Leu Leu Val Leu Leu Leu Val Thr Arg Ser Pro Val Asn Ala Cys
 -15 -10 -5 1
 Leu Leu Thr Gly Ser Leu Phe Val Leu Leu Arg Val Phe Ser Phe Glu
 5 10 15
 50 Pro Val Pro Ser Cys Arg Ala Leu Gln Val Leu Lys Pro Arg Asp Arg
 20 25 30
 Ile Ser Ala Ile Ala His Arg Gly Gly Ser His Asp Ala Pro Glu Asn
 35 40 45
 55 Thr Leu Ala Ala Ile Arg Gln Ala Ala Lys Asn Gly Ala Thr Gly Val
 50 55 60 65
 Glu Leu Asp Ile Glu Phe Thr Ser Asp Gly Ile Pro Val Leu Met His
 70 75 80
 Asp Asn Thr Val Asp Arg Thr Thr Asp Gly Thr Gly Arg Leu Cys Asp
 85 90 95
 60 Leu Thr Phe Glu Gln Ile Arg Lys Leu Asn Pro Ala Ala Asn His Arg
 100 105 110
 Leu Arg Asn Asp Phe Pro Asp Glu Lys Ile Pro Thr Leu Met Glu Ala

115 120 125
 Val Ala Glu Cys Leu Asn His Asn Leu Thr Ile Phe Phe Asp Val Lys
 130 135 140 145
 Gly His Ala His Lys Ala Thr Glu Ala Leu Lys Lys Met Tyr Met Glu
 5 150 155 160
 Phe Pro Gln Leu Tyr Asn Asn Ser Val Val Cys Ser Phe Leu Pro Glu
 165 170 175
 Val Ile Tyr Lys Met Arg Gln Thr Asp Arg Asp Val Ile Thr Ala Leu
 180 185 190
 10 Thr His Arg Pro Trp Ser Leu Ser His Thr Gly Asp Gly Lys Pro Arg
 195 200 205
 Tyr Asp Thr Phe Trp Lys His Phe Ile Phe Val Met Met Asp Ile Leu
 210 215 220 225
 Leu Asp Trp Ser Met His Asn Ile Leu Trp Tyr Leu Cys Gly Ile Ser
 15 230 235 240
 Ala Phe Leu Met Gln Lys Asp Phe Val Ser Pro Ala Tyr Leu Lys Lys
 245 250 255
 Trp Ser Ala Lys Gly Ile Gln Val Val Gly Trp Thr Val Asn Thr Phe
 260 265 270
 20 Asp Glu Lys Ser Tyr Tyr Glu Ser His Leu Gly Ser Ser Tyr Ile Thr
 275 280 285
 Asp Ser Met Val Glu Asp Cys Glu Pro His Phe
 290 295 300

 25 <210> 244
 <211> 274
 <212> PRT
 <213> Homo sapiens

 30 <220>
 <221> SIGNAL
 <222> -17...-1

 <400> 244
 35 Met Asp Arg Pro Gly Phe Val Ala Ala Leu Val Ala Gly Gly Val Ala
 -15 -10 -5
 Gly Val Ser Val Asp Leu Ile Leu Phe Pro Leu Asp Thr Ile Lys Thr
 1 5 10 15
 Arg Leu Gln Ser Pro Gln Gly Phe Ser Lys Ala Gly Gly Phe His Gly
 40 20 25 30
 Ile Tyr Ala Gly Val Pro Ser Ala Ala Ile Gly Ser Phe Pro Asn Ala
 35 40 45
 Ala Ala Phe Phe Ile Thr Tyr Glu Tyr Val Lys Trp Phe Leu His Ala
 50 55 60
 45 Asp Ser Ser Ser Tyr Leu Thr Pro Met Lys His Met Leu Ala Ala Ser
 65 70 75
 Ala Gly Glu Val Val Ala Cys Leu Ile Arg Val Pro Ser Glu Val Val
 80 85 90 95
 Lys Gln Arg Ala Gln Val Ser Ala Ser Thr Arg Thr Phe Gln Ile Phe
 50 100 105 110
 Ser Asn Ile Leu Tyr Glu Glu Gly Ile Gln Gly Leu Tyr Arg Gly Tyr
 115 120 125
 Lys Ser Thr Val Leu Arg Glu Ile Pro Phe Ser Leu Val Gln Phe Pro
 130 135 140
 55 Leu Trp Glu Ser Leu Lys Ala Leu Trp Ser Trp Arg Gln Asp His Val
 145 150 155
 Val Asp Ser Trp Gln Ser Ala Val Cys Gly Ala Phe Ala Gly Gly Phe
 160 165 170 175
 Ala Ala Ala Val Thr Thr Pro Leu Asp Val Ala Lys Thr Arg Ile Met
 180 185 190
 60 Leu Ala Lys Ala Gly Ser Ser Thr Ala Asp Gly Asn Val Leu Ser Val
 195 200 205
 Leu His Gly Val Trp Arg Ser Gln Gly Leu Ala Gly Leu Phe Ala Gly

210 215 220
 Val Phe Pro Arg Met Ala Ala Ile Ser Leu Gly Gly Phe Ile Phe Leu
 225 230 235
 Gly Ala Tyr Asp Arg Thr His Ser Leu Leu Leu Glu Val Gly Arg Lys
 5 240 245 250 255
 Ser Pro

 <210> 245
 <211> 406
 10 <212> PRT
 <213> Homo sapiens

 <220>
 <221> SIGNAL
 15 <222> -35...-1

 <400> 245
 Met Arg Gly Ser Val Glu Cys Thr Trp Gly Trp Gly His Cys Ala Pro
 -35 -30 -25 -20
 20 Ser Pro Leu Leu Leu Trp Thr Leu Leu Leu Phe Ala Ala Pro Phe Gly
 -15 -10 -5
 Leu Leu Gly Glu Lys Thr Arg Gln Val Ser Leu Glu Val Ile Pro Asn
 1 5 10
 Trp Leu Gly Pro Leu Gln Asn Leu Leu His Ile Arg Ala Val Gly Thr
 25 15 20 25
 Asn Ser Thr Leu His Tyr Val Trp Ser Ser Leu Gly Pro Leu Ala Val
 30 35 40 45
 Val Met Val Ala Thr Asn Thr Pro His Ser Thr Leu Ser Val Asn Trp
 50 55 60
 30 Ser Leu Leu Leu Ser Pro Glu Pro Asp Gly Gly Leu Met Val Leu Pro
 65 70 75
 Lys Asp Ser Ile Gln Phe Ser Ser Ala Leu Val Phe Thr Arg Leu Leu
 80 85 90
 Glu Phe Asp Ser Thr Asn Val Ser Asp Thr Ala Ala Lys Pro Leu Gly
 35 95 100 105
 Arg Pro Tyr Pro Pro Tyr Ser Leu Ala Asp Phe Ser Trp Asn Asn Ile
 110 115 120 125
 Thr Asp Ser Leu Asp Pro Ala Thr Leu Ser Ala Thr Phe Gln Gly His
 130 135 140
 40 Pro Met Asn Asp Pro Thr Arg Thr Phe Ala Asn Gly Ser Leu Ala Phe
 145 150 155
 Arg Val Gln Ala Phe Ser Arg Ser Ser Arg Pro Ala Gln Pro Pro Arg
 160 165 170
 Leu Leu His Thr Ala Asp Thr Cys Gln Leu Glu Val Ala Leu Ile Gly
 45 175 180 185
 Ala Ser Pro Arg Gly Asn Arg Ser Leu Phe Gly Leu Glu Val Ala Thr
 190 195 200 205
 Leu Gly Gln Gly Pro Asp Cys Pro Ser Met Gln Glu Gln His Ser Ile
 210 215 220
 50 Asp Asp Glu Tyr Ala Pro Ala Val Phe Gln Leu Asp Gln Leu Leu Trp
 225 230 235
 Gly Ser Leu Pro Ser Gly Phe Ala Gln Trp Arg Pro Val Ala Tyr Ser
 240 245 250
 Gln Lys Pro Gly Gly Arg Glu Ser Ala Leu Pro Cys Gln Ala Ser Pro
 55 255 260 265
 Leu His Pro Ala Leu Ala Tyr Ser Leu Pro Gln Ser Pro Ile Val Arg
 270 275 280 285
 Ala Phe Phe Gly Ser Gln Asn Asn Phe Cys Ala Phe Asn Leu Thr Phe
 290 295 300
 60 Gly Ala Ser Thr Gly Pro Gly Tyr Trp Asp Gln His Tyr Leu Ser Trp
 305 310 315
 Ser Met Leu Leu Gly Val Gly Phe Pro Pro Val Asp Gly Leu Ser Pro
 320 325 330

Leu Val Leu Gly Ile Met Ala Val Ala Leu Gly Ala Pro Gly Leu Met
 335 340 345
 Leu Leu Gly Gly Gly Leu Val Leu Leu Leu His His Lys Lys Tyr Ser
 350 355 360 365
 5 Glu Tyr Gln Ser Ile Asn
 370

 <210> 246
 <211> 24
 10 <212> PRT
 <213> Homo sapiens

 <220>
 <221> SIGNAL
 15 <222> -16..-1

 <400> 246
 Met Ala Pro Leu Gly Met Leu Leu Gly Leu Leu Met Ala Ala Cys Thr
 -15 -10 -5
 20 Pro Ser Ala Ser Val Ile Arg Thr
 1 5

 <210> 247
 <211> 348
 25 <212> PRT
 <213> Homo sapiens

 <220>
 <221> SIGNAL
 30 <222> -29..-1

 <400> 247
 Met Ala Pro Gln Ser Leu Pro Ser Ser Arg Met Ala Pro Leu Gly Met
 -25 -20 -15
 35 Leu Leu Gly Pro Leu Met Ala Ala Cys Phe Thr Phe Cys Leu Ser His
 -10 -5 1
 Gln Asn Leu Lys Glu Phe Ala Leu Thr Asn Pro Glu Lys Ser Ser Thr
 5 10 15
 Lys Glu Thr Glu Arg Lys Glu Thr Lys Ala Glu Glu Glu Leu Asp Ala
 40 20 25 30 35
 Glu Val Leu Glu Val Phe His Pro Thr His Glu Trp Gln Ala Leu Gln
 40 45 50
 Pro Gly Gln Ala Val Pro Ala Gly Ser His Val Arg Leu Asn Leu Gln
 55 60 65
 45 Thr Gly Glu Arg Glu Ala Lys Leu Gln Tyr Glu Asp Lys Phe Arg Asn
 70 75 80
 Asn Leu Lys Gly Lys Arg Leu Asp Ile Asn Thr Asn Thr Tyr Thr Ser
 85 90 95
 Gln Asp Leu Lys Ser Ala Leu Ala Lys Phe Lys Glu Gly Ala Glu Met
 50 100 105 110 115
 Glu Ser Ser Lys Glu Asp Lys Ala Arg Gln Ala Glu Val Lys Arg Leu
 120 125 130
 Phe Arg Pro Ile Glu Glu Leu Lys Lys Asp Phe Asp Glu Leu Asn Val
 135 140 145
 55 Val Ile Glu Thr Asp Met Gln Ile Met Val Arg Leu Ile Asn Lys Phe
 150 155 160
 Asn Ser Ser Ser Ser Ser Leu Glu Glu Lys Ile Ala Ala Leu Phe Asp
 165 170 175
 Leu Glu Tyr Tyr Val His Gln Met Asp Asn Ala Gln Asp Leu Leu Ser
 60 180 185 190 195
 Phe Gly Gly Leu Gln Val Val Ile Asn Gly Leu Asn Ser Thr Glu Pro
 200 205 210
 Leu Val Lys Glu Tyr Ala Ala Phe Val Leu Gly Ala Ala Phe Ser Ser

215 220 225
 Asn Pro Lys Val Gln Val Glu Ala Ile Glu Gly Gly Ala Leu Gln Lys
 230 235 240
 Leu Leu Val Ile Leu Ala Thr Glu Gln Pro Leu Thr Ala Lys Lys Lys
 5 245 250 255
 Val Leu Phe Ala Leu Cys Ser Leu Leu Arg His Phe Pro Tyr Ala Gln
 260 265 270 275
 Arg Gln Phe Leu Lys Leu Gly Gly Leu Gln Val Leu Arg Thr Leu Val
 280 285 290
 10 Gln Glu Lys Gly Thr Glu Val Leu Ala Val Arg Val Val Thr Leu Leu
 295 300 305
 Tyr Asp Leu Val Thr Glu Lys Met Phe Ala Glu Glu
 310 315

 15 <210> 248
 <211> 397
 <212> PRT
 <213> Homo sapiens

 20 <220>
 <221> SIGNAL
 <222> -36..-1

 <400> 248
 25 Met Glu Glu Leu Gln Glu Pro Leu Arg Gly Gln Leu Arg Leu Cys Phe
 -35 -30 -25
 Thr Gln Ala Ala Arg Thr Ser Leu Leu Leu Leu Arg Leu Asn Asp Ala
 -20 -15 -10 -5
 Ala Leu Arg Ala Leu Gln Glu Cys Gln Arg Gln Gln Val Arg Pro Val
 30 1 5 10
 Ile Ala Phe Gln Gly His Arg Gly Tyr Leu Arg Leu Pro Gly Pro Gly
 15 20 25
 Trp Ser Cys Leu Phe Ser Phe Ile Val Ser Gln Cys Cys Gln Glu Gly
 30 35 40
 35 Ala Gly Gly Ser Leu Asp Leu Val Cys Gln Arg Phe Leu Arg Ser Gly
 45 50 55 60
 Pro Asn Ser Leu His Cys Leu Gly Ser Leu Arg Glu Arg Leu Ile Ile
 65 70 75
 Trp Ala Ala Met Asp Ser Ile Pro Ala Pro Ser Ser Val Gln Gly His
 40 80 85 90
 Asn Leu Thr Glu Asp Ala Arg His Pro Glu Ser Trp Gln Asn Thr Gly
 95 100 105
 Gly Tyr Ser Glu Gly Asp Ala Val Ser Gln Pro Gln Met Ala Leu Glu
 110 115 120
 45 Glu Val Ser Val Ser Asp Pro Leu Ala Ser Asn Gln Gly Gln Ser Leu
 125 130 135 140
 Pro Gly Ser Ser Arg Glu His Met Ala Gln Trp Glu Val Arg Ser Gln
 145 150 155
 Thr His Val Pro Asn Arg Glu Pro Val Gln Ala Leu Pro Ser Ser Ala
 50 160 165 170
 Ser Arg Lys Arg Leu Asp Lys Lys Arg Ser Val Pro Val Ala Thr Val
 175 180 185
 Glu Leu Glu Glu Lys Arg Phe Arg Thr Leu Pro Leu Val Pro Ser Pro
 190 195 200
 55 Leu Gln Gly Leu Thr Asn Gln Asp Leu Gln Glu Gly Glu Asp Trp Glu
 205 210 215 220
 Gln Glu Asp Glu Asp Met Asp Pro Arg Leu Glu His Ser Ser Ser Val
 225 230 235
 Gln Glu Asp Ser Glu Ser Pro Ser Pro Glu Asp Ile Pro Asp Tyr Leu
 60 240 245 250
 Leu Gln Tyr Arg Ala Ile His Ser Ala Glu Gln Gln His Ala Tyr Glu
 255 260 265
 Gln Asp Phe Glu Thr Asp Tyr Ala Glu Tyr Arg Ile Leu His Ala Arg

270 275 280
 Val Gly Thr Ala Ser Gln Arg Phe Ile Glu Leu Gly Ala Glu Ile Lys
 285 290 295 300
 Arg Val Arg Arg Gly Thr Pro Glu Tyr Lys Val Leu Glu Asp Lys Ile
 5 305 310 315
 Ile Gln Glu Tyr Lys Lys Phe Arg Lys Gln Tyr Pro Ser Tyr Arg Glu
 320 325 330
 Glu Lys Arg Arg Cys Glu Tyr Leu His Gln Lys Leu Ser His Ile Lys
 335 340 345
 10 Gly Leu Ile Leu Glu Phe Glu Glu Lys Asn Arg Gly Ser
 350 355 360

 <210> 249
 <211> 403
 15 <212> PRT
 <213> Homo sapiens

 <220>
 <221> SIGNAL
 20 <222> -21...-1

 <400> 249
 Met Val Asn Asp Pro Pro Val Pro Ala Leu Leu Trp Ala Gln Glu Val
 -20 -15 -10
 25 Gly Gln Val Leu Ala Gly Arg Ala Arg Arg Leu Leu Leu Gln Phe Gly
 -5 1 5 10
 Val Leu Phe Cys Thr Ile Leu Leu Leu Trp Val Ser Val Phe Leu
 15 20 25
 Tyr Gly Ser Phe Tyr Tyr Ser Tyr Met Pro Thr Val Ser His Leu Ser
 30 30 35 40
 Pro Val His Phe Tyr Tyr Arg Thr Asp Cys Asp Ser Ser Thr Thr Ser
 45 50 55
 Leu Cys Ser Phe Pro Val Ala Asn Val Ser Leu Thr Lys Gly Gly Arg
 60 65 70 75
 35 Asp Arg Val Leu Met Tyr Gly Gln Pro Tyr Arg Val Thr Leu Glu Leu
 80 85 90
 Glu Leu Pro Glu Ser Pro Val Asn Gln Asp Leu Gly Met Phe Leu Val
 95 100 105
 Thr Ile Ser Cys Tyr Thr Arg Gly Gly Arg Ile Ile Ser Thr Ser Ser
 40 110 115 120
 Arg Ser Val Met Leu His Tyr Arg Ser Asp Leu Leu Gln Met Leu Asp
 125 130 135
 Thr Leu Val Phe Ser Ser Leu Leu Leu Phe Gly Phe Ala Glu Gln Lys
 140 145 150 155
 45 Gln Leu Leu Glu Val Glu Leu Tyr Ala Asp Tyr Arg Glu Asn Ser Val
 160 165 170
 Ser Glu Tyr Val Pro Thr Thr Gly Ala Ile Ile Glu Ile His Ser Lys
 175 180 185
 Arg Ile Gln Leu Tyr Gly Ala Tyr Leu Arg Ile His Ala His Phe Thr
 50 190 195 200
 Gly Leu Arg Tyr Leu Leu Tyr Asn Phe Pro Met Thr Cys Ala Phe Ile
 205 210 215
 Gly Val Ala Ser Asn Phe Thr Phe Leu Ser Val Ile Val Leu Phe Ser
 220 225 230 235
 55 Tyr Met Gln Trp Val Trp Gly Gly Ile Trp Pro Arg His Arg Phe Ser
 240 245 250
 Leu Gln Val Asn Ile Arg Lys Arg Asp Asn Ser Arg Lys Glu Val Gln
 255 260 265
 Arg Arg Ile Ser Ala His Gln Pro Gly Ala Gly Pro Glu Gly Gln Glu
 60 270 275 280
 Glu Ser Thr Pro Gln Ser Asp Val Thr Glu Asp Gly Glu Ser Pro Glu
 285 290 295
 Asp Pro Ser Gly Thr Glu Gly Gln Leu Ser Glu Glu Lys Pro Asp

300 305 310 315
 Gln Gln Pro Leu Ser Gly Glu Glu Glu Leu Glu Pro Glu Ala Ser Asp
 320 325 330
 Gly Ser Gly Ser Trp Glu Asp Ala Ala Leu Leu Thr Glu Ala Asn Leu
 5 335 340 345
 Pro Ala Pro Ala Pro Ala Ser Ala Ser Ala Pro Val Leu Glu Thr Leu
 350 355 360
 Gly Ser Ser Glu Pro Ala Gly Gly Ala Leu Arg Gln Arg Pro Thr Cys
 365 370 375
 10 Ser Ser Ser
 380

<210> 250
 <211> 111
 15 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 20 <222> -26..-1

<400> 250
 Met Pro His Leu Met Glu Arg Met Val Gly Ser Gly Leu Leu Trp Leu
 -25 -20 -15
 25 Ala Leu Val Ser Cys Ile Leu Thr Gln Ala Ser Ala Val Gln Arg Gly
 -10 -5 1 5
 Tyr Gly Asn Pro Ile Glu Ala Ser Ser Tyr Gly Leu Asp Leu Asp Cys
 10 15 20
 Gly Ala Pro Gly Thr Pro Glu Ala His Val Cys Phe Asp Pro Cys Gln
 30 25 30 35
 Asn Tyr Thr Leu Leu Asp Leu Gly Pro Ile Thr Arg Arg Gly Ala Gln
 40 45 50
 Ser Pro Gly Val Met Asn Gly Thr Pro Ser Thr Ala Gly Phe Leu Val
 55 60 65 70
 35 Ala Trp Pro Met Val Leu Leu Thr Val Leu Leu Ala Trp Leu Phe
 75 80 85

<210> 251
 <211> 72
 40 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 45 <222> -17..-1

<400> 251
 Met Asp Arg Pro Gly Phe Val Ala Ala Leu Val Ala Gly Gly Val Ala
 -15 -10 -5
 50 Gly Val Ser Val Asp Leu Ile Leu Phe Pro Leu Asp Thr Ile Lys Thr
 1 5 10 15
 Arg Leu Gln Ser Pro Gln Gly Phe Asn Lys Ala Gly Gly Phe His Gly
 20 25 30
 Ile Tyr Ala Gly Val Pro Ser Ala Ala Ile Gly Ser Phe Pro Asn Gly
 55 35 40 45
 Cys Leu Pro Asp Ser Ser Ser Ile
 50 55

<210> 252
 60 <211> 138
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -15...-1

5 <400> 252
 Met Lys Phe Thr Thr Leu Leu Phe Leu Ala Ala Val Ala Gly Ala Leu
 -15 -10 -5 1
 Val Tyr Ala Glu Asp Ala Ser Ser Asp Ser Thr Gly Ala Asp Pro Ala
 5 10 15
 10 Gln Glu Ala Gly Thr Ser Lys Pro Asn Glu Glu Ile Ser Gly Pro Ala
 20 25 30
 Glu Pro Ala Ser Pro Pro Glu Thr Thr Thr Thr Ala Gln Glu Thr Ser
 35 40 45
 Ala Ala Ala Val Gln Gly Thr Ala Lys Val Thr Ser Ser Arg Gln Glu
 15 50 55 60 65
 Leu Asn Pro Leu Lys Ser Ile Val Glu Lys Ser Ile Leu Leu Thr Glu
 70 75 80
 Gln Ala Leu Ala Lys Ala Gly Lys Gly Met His Gly Gly Val Pro Gly
 85 90 95
 20 Gly Lys Gln Phe Ile Glu Asn Gly Ser Glu Phe Ala Gln Lys Leu Leu
 100 105 110
 Lys Lys Phe Ser Leu Leu Lys Pro Trp Ala
 115 120

25 <210> 253
 <211> 108
 <212> PRT
 <213> Homo sapiens

30 <220>
 <221> SIGNAL
 <222> -31...-1

<220>
 35 <221> UNSURE
 <222> 45
 <223> Xaa = Glu,Gln

<220>
 40 <221> UNSURE
 <222> 44
 <223> Xaa = Lys,Asn

<400> 253
 45 Met Trp Leu Trp Glu Asp Gln Gly Gly Leu Leu Gly Pro Phe Ser Phe
 -30 -25 -20
 Leu Leu Leu Val Leu Leu Leu Val Thr Arg Ser Pro Val Asn Ala Cys
 -15 -10 -5 1
 Leu Leu Thr Gly Ser Leu Phe Val Leu Leu Arg Val Phe Ser Phe Glu
 5 10 15
 50 Pro Val Pro Ser Cys Arg Ala Leu Gln Val Leu Lys Pro Arg Asp Arg
 20 25 30
 Ile Ser Ala Ile Ala His Arg Gly Gly Ser Xaa Xaa Ala Pro Glu Asn
 35 40 45
 55 Thr Leu Ala Ala Ile Arg Gln Leu Arg Met Glu Gln Gln Ala Trp Ser
 50 55 60 65
 Trp Thr Leu Ser Leu Leu Leu Thr Gly Phe Leu Ser
 70 75

60 <210> 254
 <211> 147
 <212> PRT
 <213> Homo sapiens

```

5
<220>
<221> SIGNAL
<222> -24...-1

<400> 254
Met Val Met Gly Leu Gly Val Leu Leu Leu Val Phe Val Leu Gly Leu
-20 -15 -10
10 Gly Leu Thr Pro Pro Thr Leu Ala Gln Asp Asn Ser Arg Tyr Thr His
-5 1 5
Phe Leu Thr Gln His Tyr Asp Ala Lys Pro Gln Gly Arg Asp Asp Arg
10 15 20
Tyr Cys Glu Ser Ile Met Arg Arg Arg Gly Leu Thr Ser Pro Cys Lys
25 30 35 40
15 Asp Ile Asn Thr Phe Ile His Gly Asn Lys Arg Thr Ile Lys Ala Ile
45 50 55
Cys Glu Asn Lys Asn Gly Asn Pro His Arg Glu Asn Leu Arg Ile Ser
60 65 70
20 Lys Ser Ser Phe Gln Val Thr Thr Cys Lys Leu His Gly Gly Ser Pro
75 80 85
Trp Pro Pro Cys Gln Tyr Arg Ala Thr Ala Gly Phe Arg Asn Val Val
90 95 100
Val Ala Cys Glu Asn Gly Leu Pro Val His Leu Asp Gln Ser Ile Phe
105 110 115 120
25 Arg Arg Pro

<210> 255
<211> 381
<212> PRT
30 <213> Homo sapiens

<220>
<221> SIGNAL
<222> -33...-1

35
<400> 255
Met Ser Trp Thr Val Pro Val Val Arg Ala Ser Gln Arg Val Ser Ser
-30 -25 -20
40 Val Gly Ala Asn Phe Leu Cys Leu Gly Met Ala Leu Cys Pro Arg Gln
-15 -10 -5
Ala Thr Arg Ile Pro Leu Asn Gly Thr Trp Leu Phe Thr Pro Val Ser
1 5 10 15
Lys Met Ala Thr Val Lys Ser Glu Leu Ile Glu Arg Phe Thr Ser Glu
20 25 30
45 Lys Pro Val His Ser Lys Val Ser Ile Ile Gly Thr Gly Ser Val
35 40 45
Gly Met Ala Cys Ala Ile Ser Ile Leu Leu Lys Gly Leu Ser Asp Glu
50 55 60
Leu Ala Leu Val Asp Leu Asp Glu Asp Lys Leu Lys Gly Glu Thr Met
65 70 75
50 Asp Leu Gln His Gly Ser Pro Phe Thr Lys Met Pro Asn Ile Val Cys
80 85 90 95
Ser Lys Asp Tyr Phe Val Thr Ala Asn Ser Asn Leu Val Ile Ile Thr
100 105 110
55 Ala Gly Ala Arg Gln Glu Lys Gly Glu Thr Arg Leu Asn Leu Val Gln
115 120 125
Arg Asn Val Ala Ile Phe Lys Leu Met Ile Ser Ser Ile Val Gln Tyr
130 135 140
60 Ser Pro His Cys Lys Leu Ile Ile Val Ser Asn Pro Val Asp Ile Leu
145 150 155
Thr Tyr Val Ala Trp Lys Leu Ser Ala Phe Pro Lys Asn Arg Ile Ile
160 165 170 175
Gly Ser Gly Cys Asn Leu Asp Thr Ala Arg Phe Arg Phe Leu Ile Gly

```

180 185 190
 Gln Lys Leu Gly Ile His Ser Glu Ser Cys His Gly Trp Ile Leu Gly
 195 200 205
 5 Glu His Gly Asp Ser Ser Val Pro Val Trp Ser Gly Val Asn Ile Ala
 210 215 220
 Gly Val Pro Leu Lys Asp Leu Asn Ser Asp Ile Gly Thr Asp Lys Asp
 225 230 235
 Pro Glu Gln Trp Lys Asn Val His Lys Glu Val Thr Ala Thr Ala Tyr
 240 245 250 255
 10 Glu Ile Ile Lys Met Lys Gly Tyr Thr Ser Trp Ala Ile Gly Leu Ser
 260 265 270
 Val Ala Asp Leu Thr Glu Ser Ile Leu Lys Asn Leu Arg Arg Ile His
 275 280 285
 Pro Val Ser Thr Ile Ile Lys Gly Leu Tyr Gly Ile Asp Glu Glu Val
 290 295 300
 15 Phe Leu Ser Ile Pro Cys Ile Leu Gly Glu Asn Gly Ile Thr Asn Leu
 305 310 315
 Ile Lys Ile Lys Leu Thr Pro Glu Glu Glu Ala His Leu Lys Lys Ser
 320 325 330 335
 20 Ala Lys Thr Leu Trp Glu Ile Gln Asn Lys Leu Lys Leu
 340 345

<210> 256
 <211> 139
 25 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 30 <222> -33...-1

<400> 256
 Met Ser Trp Thr Val Pro Val Val Arg Ala Ser Gln Arg Met Ser Ser
 -30 -25 -20
 35 Val Gly Ala Asn Phe Leu Cys Leu Gly Met Ala Leu Cys Leu Arg Gln
 -15 -10 -5
 Ala Thr Arg Ile Pro Leu Asn Gly Thr Trp Leu Phe Thr Pro Val Ser
 1 5 10 15
 Lys Met Ala Thr Val Lys Ser Glu Leu Ile Glu Arg Phe Thr Ser Glu
 20 25 30
 40 Lys Pro Val His His Ser Lys Val Ser Ile Ile Gly Thr Gly Ser Val
 35 40 45
 Gly Met Ala Cys Ala Ile Ser Ile Leu Leu Lys Gly Leu Ser Asp Glu
 50 55 60
 45 Leu Ala Leu Val Asp Leu Asp Glu Asp Lys Leu Lys Gly Glu Thr Met
 65 70 75
 Asp Leu Gln His Gly Ser Pro Phe Thr Lys Met Pro Ile Leu Phe Val
 80 85 90 95
 Ala Lys Ile Thr Leu Ser Gln Gln Thr Pro Thr
 50 100 105

<210> 257
 <211> 265
 <212> PRT
 55 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -14...-1

60 <400> 257
 Met Asn Phe Ile Leu Phe Ile Phe Ile Pro Gly Val Phe Ser Leu Lys
 -10 -5 1

Ser Ser Thr Leu Lys Pro Thr Ile Glu Ala Leu Pro Asn Val Leu Pro
 5 10 15
 Leu Asn Glu Asp Val Asn Lys Gln Glu Glu Lys Asn Glu Asp His Thr
 20 25 30
 5 Pro Asn Tyr Ala Pro Ala Asn Glu Lys Asn Gly Asn Tyr Tyr Lys Asp
 35 40 45 50
 Ile Lys Gln Tyr Val Phe Thr Thr Gln Asn Pro Asn Gly Thr Glu Ser
 55 60 65
 10 Glu Ile Ser Val Arg Ala Thr Thr Asp Leu Asn Phe Ala Leu Lys Asn
 70 75 80
 Gly Ser Thr Pro Asn Val Pro Ala Phe Trp Thr Met Leu Ala Lys Ala
 85 90 95
 Ile Asn Gly Thr Ala Val Val Met Asp Asp Lys Asp Gln Leu Phe His
 100 105 110
 15 Pro Ile Pro Glu Ser Asp Val Asn Ala Thr Gln Gly Glu Asn Gln Pro
 115 120 125 130
 Asp Leu Glu Asp Leu Lys Ile Lys Ile Met Leu Gly Ile Ser Leu Met
 135 140 145
 20 Thr Leu Leu Leu Phe Val Val Leu Leu Ala Phe Cys Ser Ala Thr Leu
 150 155 160
 Tyr Lys Leu Arg His Leu Ser Tyr Lys Ser Cys Glu Ser Gln Tyr Ser
 165 170 175
 Val Asn Pro Glu Leu Ala Thr Met Ser Tyr Phe His Pro Ser Glu Gly
 180 185 190
 25 Val Ser Asp Thr Ser Phe Ser Lys Ser Ala Glu Ser Ser Thr Phe Leu
 195 200 205 210
 Gly Thr Thr Ser Ser Asp Met Arg Arg Ser Gly Thr Arg Thr Ser Glu
 215 220 225
 30 Ser Lys Ile Met Thr Asp Ile Ile Ser Ile Gly Ser Asp Asn Glu Met
 230 235 240
 His Glu Asn Asp Glu Ser Val Thr Arg
 245 250

 <210> 258
 35 <211> 200
 <212> PRT
 <213> Homo sapiens

 <220>
 40 <221> SIGNAL
 <222> -20..-1

 <400> 258
 45 Met Asp Ser Ser Thr Ala His Ser Pro Val Phe Leu Val Phe Pro Pro
 -20 -15 -10 -5
 Glu Ile Thr Ala Ser Glu Tyr Glu Ser Thr Glu Leu Ser Ala Thr Thr
 1 5 10
 Phe Ser Thr Gln Ser Pro Leu Gln Lys Leu Phe Ala Arg Lys Met Lys
 15 20 25
 50 Ile Leu Gly Thr Ile Gln Ile Leu Phe Gly Ile Met Thr Phe Ser Phe
 30 35 40
 Gly Val Ile Phe Leu Phe Thr Leu Leu Lys Pro Tyr Pro Arg Phe Pro
 45 50 55 60
 Phe Ile Phe Leu Ser Gly Tyr Pro Phe Trp Gly Ser Val Leu Phe Ile
 55 65 70 75
 Asn Ser Gly Ala Phe Leu Ile Ala Val Lys Arg Lys Thr Thr Glu Thr
 80 85 90
 Leu Ile Ile Leu Ser Arg Ile Met Asn Phe Leu Ser Ala Leu Gly Ala
 95 100 105
 60 Ile Ala Gly Ile Ile Leu Leu Thr Phe Gly Phe Ile Leu Asp Gln Asn
 110 115 120
 Tyr Ile Cys Gly Tyr Ser His Gln Asn Ser Gln Cys Lys Ala Val Thr
 125 130 135 140

Val Leu Phe Leu Gly Ile Leu Ile Thr Leu Met Thr Phe Ser Ile Ile
 145 150 155
 Glu Leu Phe Ile Ser Leu Pro Phe Ser Ile Leu Gly Cys His Ser Glu
 160 165 170
 5 Asp Cys Asp Cys Glu Gln Cys Cys
 175 180

 <210> 259
 <211> 394
 10 <212> PRT
 <213> Homo sapiens

 <220>
 <221> SIGNAL
 15 <222> -39...-1

 <400> 259
 Met Ala Thr Ala Gln Leu Gln Arg Thr Pro Met Ser Ala Leu Val Phe
 -35 -30 -25
 20 Pro Asn Lys Ile Ser Thr Glu His Gln Ser Leu Val Leu Val Lys Arg
 -20 -15 -10
 Leu Leu Ala Val Ser Val Ser Cys Ile Thr Tyr Leu Arg Gly Ile Phe
 -5 1 5
 25 Pro Glu Cys Ala Tyr Gly Thr Arg Tyr Leu Asp Asp Leu Cys Val Lys
 10 15 20 25
 Ile Leu Arg Glu Asp Lys Asn Cys Pro Gly Ser Thr Gln Leu Val Lys
 30 35 40
 Trp Ile Leu Gly Cys Tyr Asp Ala Leu Gln Lys Lys Tyr Leu Arg Met
 45 50 55
 30 Val Val Leu Ala Val Tyr Thr Asn Pro Glu Asp Pro Gln Thr Ile Ser
 60 65 70
 Glu Cys Tyr Gln Phe Lys Phe Lys Tyr Thr Asn Asn Gly Pro Leu Met
 75 80 85
 35 Asp Phe Ile Ser Lys Asn Gln Ser Asn Glu Ser Ser Met Leu Ser Thr
 90 95 100 105
 Asp Thr Lys Lys Ala Ser Ile Leu Leu Ile Arg Lys Ile Tyr Ile Leu
 110 115 120
 Met Gln Asn Leu Gly Pro Leu Pro Asn Asp Val Cys Leu Thr Met Lys
 125 130 135
 40 Leu Phe Tyr Tyr Asp Glu Val Thr Pro Pro Asp Tyr Gln Pro Pro Gly
 140 145 150
 Phe Lys Asp Gly Asp Cys Glu Gly Val Ile Phe Glu Gly Glu Pro Met
 155 160 165
 Tyr Leu Asn Val Gly Glu Val Ser Thr Pro Phe His Ile Phe Lys Val
 170 175 180 185
 45 Lys Val Thr Thr Glu Arg Glu Arg Met Glu Asn Ile Asp Ser Thr Ile
 190 195 200
 Leu Ser Pro Lys Gln Ile Lys Thr Pro Phe Gln Lys Ile Leu Arg Asp
 205 210 215
 50 Lys Asp Val Glu Asp Glu Gln Glu His Tyr Thr Ser Asp Asp Leu Asp
 220 225 230
 Ile Glu Thr Lys Met Glu Glu Gln Glu Lys Asn Pro Ala Ser Ser Glu
 235 240 245
 Leu Glu Glu Pro Ser Leu Val Cys Glu Glu Asp Glu Ile Met Arg Ser
 250 255 260 265
 55 Lys Glu Ser Pro Asp Leu Ser Ile Ser His Ser Gln Val Glu Gln Leu
 270 275 280
 Val Asn Lys Thr Ser Glu Leu Asp Met Ser Glu Ser Lys Thr Arg Ser
 285 290 295
 60 Gly Lys Val Phe Gln Asn Lys Met Ala Asn Gly Asn Gln Pro Val Lys
 300 305 310
 Ser Ser Lys Glu Asn Arg Lys Arg Ser Gln His Glu Ser Gly Arg Ile
 315 320 325

Val Leu His His Phe Asp Ser Ser Ser Gln Glu Ser Val Pro Lys Arg
 330 335 340 345
 Arg Lys Phe Ser Glu Pro Lys Glu His Ile
 350 355

5

<210> 260
 <211> 158
 <212> PRT
 <213> Homo sapiens

10

<220>
 <221> SIGNAL
 <222> -17...-1

15

<400> 260
 Met Ala Leu Glu Val Leu Met Leu Leu Ala Val Leu Ile Trp Thr Gly
 -15 -10 -5
 Ala Glu Asn Leu His Val Lys Ile Ser Cys Ser Leu Asp Trp Leu Met
 1 5 10 15
 20 Val Ser Val Ile Pro Val Ala Glu Ser Arg Asn Leu Tyr Ile Phe Ala
 20 25 30
 Asp Glu Leu His Leu Gly Met Gly Cys Pro Ala Asn Arg Ile His Thr
 35 40 45
 Tyr Val Tyr Glu Phe Ile Tyr Leu Val Arg Asp Cys Gly Ile Arg Thr
 50 55 60
 25 Arg Val Val Ser Glu Glu Thr Leu Leu Phe Gln Thr Glu Leu Tyr Phe
 65 70 75
 Thr Pro Arg Asn Ile Asp His Asp Pro Gln Glu Ile His Leu Glu Cys
 80 85 90 95
 30 Ser Thr Ser Arg Lys Ser Val Trp Leu Thr Pro Val Ser Thr Glu Asn
 100 105 110
 Glu Ile Lys Leu Asp Pro Ser Pro Phe Ile Ala Asp Phe Gln Thr Thr
 115 120 125
 Ala Glu Glu Leu Gly Leu Leu Ser Ser Pro Asn Leu Leu
 130 135 140

35

<210> 261
 <211> 233
 <212> PRT
 40 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -32...-1

45

<400> 261
 Met Ala Thr Pro Pro Phe Arg Leu Ile Arg Lys Met Phe Ser Phe Lys
 -30 -25 -20
 Val Ser Arg Trp Met Gly Leu Ala Cys Phe Arg Ser Leu Ala Ala Ser
 -15 -10 -5
 50 Ser Pro Ser Ile Arg Gln Lys Lys Leu Met His Lys Leu Gln Glu Glu
 1 5 10 15
 Lys Ala Phe Arg Glu Glu Met Lys Ile Phe Arg Glu Lys Ile Glu Asp
 20 25 30
 55 Phe Arg Glu Glu Met Trp Thr Phe Arg Gly Lys Ile His Ala Phe Arg
 35 40 45
 Gly Gln Ile Leu Gly Phe Trp Glu Glu Glu Arg Pro Phe Trp Glu Glu
 50 55 60
 Glu Lys Thr Phe Trp Lys Glu Glu Lys Ser Phe Trp Glu Met Glu Lys
 60 65 70 75 80
 Ser Phe Arg Glu Glu Lys Thr Phe Trp Lys Lys Tyr Arg Thr Phe
 85 90 95
 Trp Lys Glu Asp Lys Ala Phe Trp Lys Glu Asp Asn Ala Leu Trp Glu

```

      100      105      110
Arg Asp Arg Asn Leu Leu Gln Glu Asp Lys Ala Leu Trp Glu Glu Glu
      115      120      125
Lys Ala Leu Trp Val Glu Glu Arg Ala Leu Leu Glu Gly Glu Lys Ala
5      130      135      140
Leu Trp Glu Asp Lys Thr Ser Leu Trp Glu Glu Glu Asn Ala Leu Trp
145      150      155      160
Glu Glu Glu Arg Ala Phe Trp Met Glu Asn Asn Gly His Ile Ala Gly
      165      170      175
10 Glu Gln Met Leu Glu Asp Gly Pro His Asn Ala Asn Arg Gly Gln Arg
      180      185      190
Leu Leu Ala Phe Ser Arg Gly Arg Ala
      195      200

```

```

15 <210> 262
    <211> 67
    <212> PRT
    <213> Homo sapiens

```

```

20 <220>
    <221> SIGNAL
    <222> -20...-1

```

```

      <400> 262
25 Met Asp Ser Ser Thr Ala His Ser Pro Val Phe Leu Val Phe Pro Pro
   -20      -15      -10      -5
   Glu Ile Thr Ala Ser Glu Tyr Glu Ser Thr Glu Leu Ser Ala Thr Thr
      1      5      10
   Phe Ser Thr Gln Ser Pro Leu Gln Lys Leu Phe Ala Arg Lys Met Lys
30      15      20      25
   Ile Leu Gly Asp Ile His Ser Gly Ala Leu Phe Cys Ser Leu Ile Leu
      30      35      40
   Glu Pro Ser
   45

```

```

35 <210> 263
    <211> 94
    <212> PRT
    <213> Homo sapiens

```

```

40 <220>
    <221> SIGNAL
    <222> -25...-1

```

```

      <400> 263
45 Met Cys Phe Leu Val Ser Phe Asn Leu Pro Ile His Ile Ser Leu Ser
   -25      -20      -15      -10
   His Leu Phe Leu Asp Leu Ser Arg Ser Leu Trp Phe Leu Ala Cys Pro
      -5      1      5
50 Gly Leu Asn Leu Val Tyr Leu Ala Leu Asp Ser Phe Ser Asp Leu Arg
      10      15      20
   Pro Ser Leu Asn Leu Leu Phe Tyr Phe Val Pro Gly Phe Gly Val Ser
      25      30      35
   Lys Tyr Leu Thr Ser Ala Gln Pro Val Leu Gly Phe Leu Leu Leu Pro
55      40      45      50      55
   Asp Ile Asp Asn Pro Ala Leu Leu Gly Thr Glu Arg Trp Ser
      60      65

```

```

60 <210> 264
    <211> 174
    <212> PRT
    <213> Homo sapiens

```

```
<220>
<221> SIGNAL
<222> -19..-1
```

5	<400> 264															
	Met	Phe	Leu	Thr	Val	Lys	Leu	Leu	Leu	Gly	Gln	Arg	Cys	Ser	Leu	Lys
					-15					-10					-5	
	Val	Ser	Gly	Gln	Glu	Ser	Val	Ala	Thr	Leu	Lys	Arg	Leu	Val	Ser	Arg
				1				5					10			
10	Arg	Leu	Lys	Val	Pro	Glu	Glu	Gln	Gln	His	Leu	Leu	Phe	Arg	Gly	Gln
		15					20					25				
	Leu	Leu	Glu	Asp	Asp	Lys	His	Leu	Ser	Asp	Tyr	Cys	Ile	Gly	Pro	Asn
	30					35					40					45
	Ala	Ser	Ile	Asn	Val	Ile	Met	Gln	Pro	Leu	Glu	Lys	Met	Ala	Leu	Lys
15					50					55					60	
	Glu	Ala	His	Gln	Pro	Gln	Thr	Gln	Pro	Leu	Trp	His	Gln	Leu	Gly	Leu
				65				70						75		
	Val	Leu	Ala	Lys	His	Phe	Glu	Pro	Gln	Asp	Ala	Lys	Ala	Val	Leu	Gln
			80					85					90			
20	Leu	Leu	Arg	Gln	Glu	His	Glu	Glu	Arg	Leu	Gln	Lys	Ile	Ser	Leu	Glu
		95					100					105				
	His	Leu	Glu	Gln	Leu	Ala	Gln	Tyr	Leu	Leu	Ala	Glu	Glu	Pro	His	Val
	110					115					120					125
	Glu	Pro	Ala	Gly	Glu	Arg	Glu	Leu	Glu	Ala	Lys	Ala	Arg	Pro	Gln	Ser
25					130					135					140	
	Ser	Cys	Asp	Met	Glu	Glu	Lys	Glu	Glu	Ala	Ala	Ala	Asp	Gln		
				145				150					155			

```

30  <210> 265
    <211> 106
    <212> PRT
    <213> Homo sapiens

```

```

    <220>
35  <221> SIGNAL
    <222> -17..-1

```

[illegible]

```

55  <210> 266
    <211> 124
    <212> PRT
    <213> Homo sapiens

```

```

    <220>
60  <221>  SIGNAL
    <222>  -18..-1

```

<400> 266

Met Val Leu Cys Trp Leu Leu Leu Leu Val Met Ala Leu Pro Pro Gly
 -15 -10 -5
 Thr Thr Gly Val Lys Asp Cys Val Phe Cys Glu Leu Thr Asp Ser Met
 1 5 10
 5 Gln Cys Pro Gly Thr Tyr Met His Cys Gly Asp Asp Glu Asp Cys Phe
 15 20 25 30
 Thr Gly His Gly Val Ala Pro Gly Thr Gly Pro Val Ile Asn Lys Gly
 35 40 45
 Cys Leu Arg Ala Thr Ser Cys Gly Leu Glu Glu Pro Val Ser Tyr Arg
 50 55 60
 10 Gly Val Thr Tyr Ser Leu Thr Thr Asn Cys Cys Thr Gly Arg Leu Cys
 65 70 75
 Asn Arg Ala Pro Ser Ser Gln Thr Val Gly Ala Thr Thr Ser Leu Ala
 80 85 90
 15 Leu Gly Leu Gly Met Leu Leu Pro Pro Arg Leu Leu
 95 100 105

<210> 267

<211> 261

20 <212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

25 <222> -16...-1

<400> 267

Met Glu Asn Phe Ser Leu Leu Ser Ile Ser Gly Pro Pro Ile Ser Ser
 -15 -10 -5
 30 Ser Ala Leu Ser Ala Phe Pro Asp Ile Met Phe Ser Arg Ala Thr Ser
 1 5 10 15
 Leu Pro Asp Ile Ala Lys Thr Ala Val Pro Thr Glu Ala Ser Ser Pro
 20 25 30
 Ala Gln Ala Leu Pro Pro Gln Tyr Gln Ser Ile Ile Val Arg Gln Gly
 35 35 40 45
 Ile Gln Asn Thr Val Leu Ser Pro Asp Cys Ser Leu Gly Asp Thr Gln
 50 55 60
 His Gly Glu Lys Leu Arg Arg Asn Cys Thr Ile Tyr Arg Pro Trp Phe
 65 70 75 80
 40 Ser Pro Tyr Ser Tyr Phe Val Cys Ala Asp Lys Glu Ser Gln Leu Glu
 85 90 95
 Ala Tyr Asp Phe Pro Glu Val Gln Gln Asp Glu Gly Lys Trp Asp Asn
 100 105 110
 Cys Leu Ser Glu Asp Met Ala Glu Asn Ile Cys Ser Ser Ser Ser
 115 120 125
 45 Pro Glu Asn Thr Cys Pro Arg Glu Ala Thr Lys Lys Ser Arg His Gly
 130 135 140
 Leu Asp Ser Ile Thr Ser Gln Asp Ile Leu Met Ala Ser Arg Trp His
 145 150 155 160
 50 Pro Ala Gln Gln Asn Gly Tyr Lys Cys Val Ala Cys Cys Arg Met Tyr
 165 170 175
 Pro Thr Leu Asp Phe Leu Lys Ser His Ile Lys Arg Gly Phe Arg Glu
 180 185 190
 Gly Phe Ser Cys Lys Val Tyr Tyr Arg Lys Leu Lys Ala Leu Trp Ser
 195 200 205
 55 Lys Glu Gln Lys Ala Arg Leu Gly Asp Arg Leu Ser Ser Gly Ser Cys
 210 215 220
 Gln Ala Phe Asn Ser Pro Ala Glu His Leu Arg Gln Ile Gly Gly Glu
 225 230 235 240
 60 Ala Tyr Leu Cys Leu
 245

<210> 268

<211> 76
 <212> PRT
 <213> Homo sapiens

5 <220>
 <221> SIGNAL
 <222> -25...-1

<400> 268

10 Met Cys Met Ser Leu Ser Met Lys Val Pro Cys Cys Leu Cys Ala Leu
 -25 -20 -15 -10
 Leu Ser Asn Phe Cys Pro Ser Thr Thr Val Lys Gly Asp Val Val Thr
 -5 1 5
 Ser Phe Phe Arg Ala Asp Tyr Asp Leu Ala Ser Arg Ser Ala Asp Gln
 15 10 15 20
 Ser Ser Gln Lys Val Lys Leu Arg Met Phe Thr Gly Arg Leu Pro Ile
 25 30 35
 Gly Pro Phe Ala Ser Val Gly Asn Ala Ala Glu Leu
 40 45 50

20 <210> 269
 <211> 199
 <212> PRT
 <213> Homo sapiens

25 <220>
 <221> SIGNAL
 <222> -16...-1

30 <400> 269

Met Glu Thr Phe Pro Leu Leu Leu Leu Ser Leu Gly Leu Val Leu Ala
 -15 -10 -5
 Glu Ala Ser Glu Ser Thr Met Lys Ile Ile Lys Glu Glu Phe Thr Asp
 1 5 10 15
 35 Glu Glu Met Gln Tyr Asp Met Ala Lys Ser Gly Gln Glu Lys Gln Thr
 20 25 30
 Ile Glu Ile Leu Met Asn Pro Ile Leu Leu Val Lys Asn Thr Ser Leu
 35 40 45
 Ser Met Ser Lys Asp Asp Met Ser Ser Thr Leu Leu Thr Phe Arg Ser
 40 50 55 60
 Leu His Tyr Asn Asp Pro Lys Gly Asn Ser Ser Gly Asn Asp Lys Glu
 65 70 75 80
 Cys Cys Asn Asp Met Thr Val Trp Arg Lys Val Ser Glu Ala Asn Gly
 85 90 95
 45 Ser Cys Lys Trp Ser Asn Asn Phe Ile Arg Ser Ser Thr Glu Val Met
 100 105 110
 Arg Arg Val His Arg Ala Pro Ser Cys Lys Phe Val Gln Asn Pro Gly
 115 120 125
 Ile Ser Cys Cys Glu Ser Leu Glu Leu Glu Asn Thr Val Cys Gln Phe
 50 130 135 140
 Thr Thr Gly Lys Gln Phe Pro Arg Cys Gln Tyr His Ser Val Thr Ser
 145 150 155 160
 Leu Glu Lys Ile Leu Thr Val Leu Thr Gly His Ser Leu Met Ser Trp
 165 170 175
 55 Leu Val Cys Gly Ser Lys Leu
 180

<210> 270
 <211> 88
 60 <212> PRT
 <213> Homo sapiens
 <220>

<221> SIGNAL
 <222> -36...-1

<400> 270
 5 Met Ala Ser Val Val Pro Val Lys Asp Lys Lys Leu Leu Glu Val Lys
 -35 -30 -25
 Leu Gly Glu Leu Pro Ser Trp Ile Leu Met Arg Asp Phe Ser Pro Ser
 -20 -15 -10 -5
 10 Gly Ile Phe Gly Ala Phe Gln Arg Gly Tyr Tyr Arg Tyr Tyr Asn Lys
 1 5 10
 Tyr Ile Asn Val Lys Lys Gly Ser Ile Ser Gly Ile Thr Met Val Leu
 15 20 25
 Ala Cys Tyr Val Leu Phe Ser Tyr Ser Phe Ser Tyr Lys His Leu Lys
 30 35 40
 15 His Glu Arg Leu Arg Lys Tyr His
 45 50

<210> 271
 <211> 481
 20 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 25 <222> -25...-1

<400> 271
 Met Gly Ala Leu Ala Arg Ala Leu Pro Ser Ile Leu Leu Ala Leu Leu
 -25 -20 -15 -10
 30 Leu Thr Ser Thr Pro Glu Ala Leu Gly Ala Asn Pro Gly Leu Val Ala
 -5 1 5
 Arg Ile Thr Asp Lys Gly Leu Gln Tyr Ala Ala Gln Glu Gly Leu Leu
 10 15 20
 Ala Leu Gln Ser Glu Leu Leu Arg Ile Thr Leu Pro Asp Phe Thr Gly
 25 30 35
 35 Asp Leu Arg Ile Pro His Val Gly Arg Gly Arg Tyr Glu Phe His Ser
 40 45 50 55
 Leu Asn Ile His Ser Cys Glu Leu Leu His Ser Ala Leu Arg Pro Val
 60 65 70
 40 Pro Gly Gln Gly Leu Ser Leu Ser Ile Ser Asp Ser Ser Ile Arg Val
 75 80 85
 Gln Gly Arg Trp Lys Val Arg Lys Ser Phe Phe Lys Leu Gln Gly Ser
 90 95 100
 Phe Asp Val Ser Val Lys Gly Ile Ser Ile Ser Val Asn Leu Leu Leu
 45 Gly Ser Asp Ser Ser Gly Arg Pro Thr Val Thr Ala Ser Ser Cys Ser
 105 110 115
 120 125 130 135
 Ser Asp Ile Ala Asp Val Glu Val Asp Met Ser Gly Asp Leu Gly Trp
 140 145 150
 50 Leu Leu Asn Leu Phe His Asn Gln Ile Glu Ser Lys Phe Gln Lys Val
 155 160 165
 Leu Glu Ser Arg Ile Cys Glu Met Ile Gln Lys Ser Val Ser Ser Asp
 170 175 180
 Leu Gln Pro Tyr Leu Gln Thr Leu Thr Val Thr Thr Glu Ile Asp Ser
 55 Phe Ala Asp Ile Asp Tyr Ser Leu Val Glu Ala Pro Arg Ala Thr Ala
 185 190 195 200
 205 210 215
 Gln Met Leu Glu Val Met Phe Lys Gly Glu Ile Phe His Arg Asn His
 220 225 230
 60 Arg Ser Pro Val Thr Leu Leu Ala Ala Val Met Ser Leu Pro Glu Glu
 235 240 245
 His Asn Lys Met Val Tyr Phe Ala Ile Ser Asp Tyr Val Phe Asn Thr
 250 255 260

Ala Ser Leu Val Tyr His Glu Glu Gly Tyr Leu Asn Phe Ser Ile Thr
 265 270 275
 Asp Asp Met Ile Pro Pro Asp Ser Asn Ile Arg Leu Thr Thr Lys Ser
 280 285 290 295
 5 Phe Arg Pro Phe Val Pro Arg Leu Ala Arg Leu Tyr Pro Asn Met Asn
 300 305 310
 Leu Glu Leu Gln Gly Ser Val Pro Ser Ala Pro Leu Leu Asn Phe Ser
 315 320 325
 10 Pro Gly Asn Leu Ser Val Asp Pro Tyr Met Glu Ile Asp Ala Phe Val
 330 335 340
 Leu Leu Pro Ser Ser Ser Lys Glu Pro Val Phe Arg Leu Ser Val Ala
 345 350 355
 Thr Asn Val Ser Ala Thr Leu Thr Phe Asn Thr Ser Lys Ile Thr Gly
 360 365 370 375
 15 Phe Leu Lys Pro Gly Lys Val Lys Val Glu Leu Lys Glu Ser Lys Val
 380 385 390
 Gly Leu Phe Asn Ala Glu Leu Leu Glu Ala Leu Leu Asn Tyr Tyr Ile
 395 400 405
 20 Leu Asn Thr Phe Tyr Pro Lys Phe Asn Asp Lys Leu Ala Glu Gly Phe
 410 415 420
 Pro Leu Pro Leu Leu Lys Arg Val Gln Leu Tyr Asp Leu Gly Leu Gln
 425 430 435
 Ile His Lys Asp Phe Leu Phe Leu Gly Ala Asn Val Gln Tyr Met Arg
 440 445 450 455
 25 Val

<210> 272

<211> 143

<212> PRT

30 <213> Homo sapiens

<220>

<221> SIGNAL

<222> -43...-1

35

<400> 272

Met Ala Lys Tyr Gln Gly Glu Val Gln Ser Leu Lys Leu Asp Asp Asp
 -40 -35 -30

40 Ser Val Ile Glu Gly Val Ser Asp Gln Val Leu Val Ala Val Val Val
 -25 -20 -15

Ser Phe Ala Leu Ile Ala Thr Leu Val Tyr Ala Leu Phe Arg Asn Val
 -10 -5 1 5

His Gln Asn Ile His Pro Glu Asn Gln Glu Leu Val Arg Val Leu Arg
 10 15 20

45 Glu Gln Leu Gln Thr Glu Gln Asp Ala Pro Ala Ala Thr Arg Gln Gln
 25 30 35

Phe Tyr Thr Asp Met Tyr Cys Pro Ile Cys Leu His Gln Ala Ser Phe
 40 45 50

50 Pro Val Glu Thr Asn Cys Gly His Leu Phe Cys Gly Ala Cys Ile Ile
 55 60 65

Ala Tyr Trp Arg Tyr Gly Ser Trp Leu Gly Ala Ile Ser Cys Pro Ile
 70 75 80 85

Cys Arg Gln Thr Arg His Gly His Ile Ala Leu Ser Arg Thr Ala
 90 95 100

55

<210> 273

<211> 82

<212> PRT

<213> Homo sapiens

60

<400> 273

Met Ala Lys Tyr Gln Gly Glu Val Gln Ser Leu Lys Leu Asp Asp Asp
 1 5 10 15

Ser Val Ile Glu Gly Val Ser Asp Gln Val Leu Val Ala Val Val Val
 20 25 30
 Ser Phe Ala Leu Ile Ala Thr Leu Val Tyr Ala Leu Phe Arg Asn Val
 35 40 45
 5 His Gln Asn Ile His Pro Glu Asn Gln Glu Leu Val Arg Val Leu Arg
 50 55 60
 Glu Gln Leu Gln Thr Glu Gln Asp Ala Pro Ala Asp Ser Thr Ala Val
 65 70 75 80
 Leu His
 10
 <210> 274
 <211> 373
 <212> PRT
 <213> Homo sapiens
 15
 <220>
 <221> SIGNAL
 <222> -27...-1
 20 <400> 274
 Met Ala Thr Gln Ala His Ser Leu Ser Tyr Ala Gly Cys Asn Phe Leu
 -25 -20 -15
 Cys Gln Arg Leu Val Leu Ser Thr Leu Ser Gly Arg Pro Val Lys Ile
 -10 -5 1 5
 25 Arg Lys Ile Arg Ala Arg Asp Asp Asn Pro Gly Leu Arg Asp Phe Glu
 10 15 20
 Ala Ser Phe Ile Arg Leu Leu Asp Lys Ile Thr Asn Gly Ser Arg Ile
 25 30 35
 Glu Ile Asn Gln Thr Gly Thr Thr Leu Tyr Tyr Gln Pro Gly Leu Leu
 40 45 50
 Tyr Gly Gly Ser Val Glu His Asp Cys Ser Val Leu Arg Gly Ile Gly
 55 60 65
 Tyr Tyr Leu Glu Ser Leu Leu Cys Leu Ala Pro Phe Met Lys His Pro
 70 75 80 85
 35 Leu Lys Ile Val Leu Arg Gly Val Thr Asn Asp Gln Ile Asp Pro Ser
 90 95 100
 Val Asp Val Leu Lys Ala Thr Ala Leu Pro Leu Leu Lys Gln Phe Gly
 105 110 115
 Ile Asp Gly Glu Ser Phe Glu Leu Lys Ile Val Arg Arg Gly Met Pro
 120 125 130
 40 Pro Gly Gly Gly Gly Glu Val Val Phe Ser Cys Pro Val Arg Lys Val
 135 140 145
 Leu Lys Pro Ile Gln Leu Thr Asp Pro Gly Lys Ile Lys Arg Ile Arg
 150 155 160 165
 45 Gly Met Ala Tyr Ser Val Arg Val Ser Pro Gln Met Ala Asn Arg Ile
 170 175 180
 Val Asp Ser Ala Arg Ser Ile Leu Asn Lys Phe Ile Pro Asp Ile Tyr
 185 190 195
 Ile Tyr Thr Asp His Ile Lys Gly Val Asn Ser Gly Lys Ser Pro Gly
 200 205 210
 50 Phe Gly Leu Ser Leu Val Ala Glu Thr Thr Ser Gly Thr Phe Leu Ser
 215 220 225
 Ala Glu Leu Ala Ser Asn Pro Gln Gly Gln Gly Ala Ala Val Leu Pro
 230 235 240 245
 55 Glu Asp Leu Gly Arg Asn Cys Ala Arg Leu Leu Glu Glu Ile Tyr
 250 255 260
 Arg Gly Gly Cys Val Asp Ser Thr Asn Gln Ser Leu Ala Leu Leu Leu
 265 270 275
 Met Thr Leu Gly Gln Gln Asp Val Ser Lys Val Leu Leu Gly Pro Leu
 280 285 290
 60 Ser Pro Tyr Thr Ile Glu Phe Leu Arg His Leu Lys Ser Phe Phe Gln
 295 300 305
 Ile Met Phe Lys Ile Glu Thr Lys Pro Cys Gly Glu Glu Leu Lys Gly

310 315 320 325
 Gly Asp Lys Val Leu Met Thr Cys Val Gly Ile Gly Phe Ser Asn Leu
 330 335 340
 Ser Arg Thr Leu Lys
 5 345

 <210> 275
 <211> 94
 <212> PRT
 10 <213> Homo sapiens

 <220>
 <221> SIGNAL
 <222> -25...-1
 15
 <400> 275
 Met Ala Ser Val Val Leu Ala Leu Arg Thr Arg Thr Ala Val Thr Ser
 -25 -20 -15 -10
 20 Leu Leu Ser Pro Thr Pro Ala Thr Ala Leu Ala Val Arg Tyr Ala Ser
 -5 1 5
 Lys Lys Ser Gly Gly Ser Ser Lys Asn Leu Gly Gly Lys Ser Ser Gly
 10 15 20
 Arg Arg Gln Gly Ile Lys Lys Met Glu Gly His Tyr Val His Ala Gly
 25 30 35
 25 Asn Ile Ile Ala Thr Gln Arg His Phe Arg Trp His Pro Gly Ala His
 40 45 50 55
 Val Ser Cys Ser Val Ala Ala Pro Leu Phe Pro Phe Leu Gly
 60 65

 30 <210> 276
 <211> 197
 <212> PRT
 <213> Homo sapiens

 35 <220>
 <221> SIGNAL
 <222> -20...-1

 <400> 276
 40 Met Thr Val Leu Glu Ile Thr Leu Ala Val Ile Leu Thr Leu Leu Gly
 -20 -15 -10 -5
 Leu Ala Ile Leu Ala Ile Leu Leu Thr Arg Trp Ala Arg Arg Lys Gln
 1 5 10
 Ser Glu Met Tyr Ile Ser Arg Tyr Ser Ser Glu Gln Ser Ala Arg Leu
 45 15 20 25
 Leu Asp Tyr Glu Asp Gly Arg Gly Ser Arg His Ala Tyr Ser Thr Gln
 30 35 40
 Ser Glu Arg Ser Lys Arg Asp Tyr Thr Pro Ser Thr Asn Ser Leu Ala
 45 50 55 60
 50 Leu Ser Arg Ser Ser Ile Ala Leu Pro Gln Gly Ser Met Ser Ser Ile
 65 70 75
 Lys Cys Leu Gln Thr Thr Glu Glu Pro Ser Arg Thr Ala Gly Ala
 80 85 90
 Met Met Gln Phe Thr Ala Pro Ile Pro Gly Ala Thr Gly Pro Ile Lys
 55 95 100 105
 Leu Ser Gln Lys Thr Ile Val Gln Thr Leu Gly Pro Ile Val Gln Tyr
 110 115 120
 Pro Gly Ser Asn Gly Arg Ile Asn Ile Ser Gln Leu Thr Ser Glu Asp
 125 130 135 140
 60 Leu Thr Gly Ala Lys Gly Arg Val Thr Ser Gly Pro Gln Phe Pro Asn
 145 150 155
 Ser His His Val Pro Glu Asn Leu His Gly Tyr Met Asn Ser Leu Ser
 160 165 170

Leu Phe Ser Pro Ala
175

<210> 277
5 <211> 344
<212> PRT
<213> Homo sapiens

<220>
10 <221> SIGNAL
<222> -29...-1

<400> 277
15 Met Asp Phe Leu Val Leu Phe Leu Phe Tyr Leu Ala Ser Val Leu Met
-25 -20 -15
Gly Leu Val Leu Ile Cys Val Cys Ser Lys Thr His Ser Leu Lys Gly
-10 -5 1
Leu Ala Arg Gly Gly Ala Gln Ile Phe Ser Cys Ile Ile Pro Glu Cys
5 10 15
20 Leu Gln Arg Ala Val His Gly Leu Leu His Tyr Leu Phe His Thr Arg
20 25 30 35
Asn His Thr Phe Ile Val Leu His Leu Val Leu Gln Gly Met Val Tyr
40 45 50
25 Thr Glu Tyr Thr Trp Glu Val Phe Gly Tyr Cys Gln Glu Leu Glu Leu
55 60 65
Ser Leu His Tyr Leu Leu Leu Pro Tyr Leu Leu Leu Gly Val Asn Leu
70 75 80
Phe Phe Phe Thr Leu Thr Cys Gly Thr Asn Pro Gly Ile Ile Thr Lys
85 90 95
30 Ala Asn Glu Leu Leu Phe Leu His Val Tyr Glu Phe Asp Glu Val Met
100 105 110 115
Phe Pro Lys Asn Val Arg Cys Ser Thr Cys Asp Leu Arg Lys Pro Ala
120 125 130
35 Arg Ser Lys His Cys Ser Val Cys Asn Trp Cys Val His Arg Phe Asp
135 140 145
His His Cys Val Trp Val Asn Asn Cys Ile Gly Ala Trp Asn Ile Arg
150 155 160
Tyr Phe Leu Ile Tyr Val Leu Thr Leu Thr Ala Ser Ala Ala Thr Val
165 170 175
40 Ala Ile Val Ser Thr Thr Phe Leu Val His Leu Val Val Met Ser Asp
180 185 190 195
Leu Tyr Gln Glu Thr Tyr Ile Asp Asp Leu Gly His Leu His Val Met
200 205 210
45 Asp Thr Val Phe Leu Ile Gln Tyr Leu Phe Leu Thr Phe Pro Arg Ile
215 220 225
Val Phe Met Leu Gly Phe Val Val Val Leu Ser Phe Leu Leu Gly Gly
230 235 240
Tyr Leu Leu Phe Val Leu Tyr Leu Ala Ala Thr Asn Gln Thr Thr Asn
245 250 255
50 Glu Trp Tyr Arg Gly Asp Trp Ala Trp Cys Gln Arg Cys Pro Leu Val
260 265 270 275
Ala Trp Pro Pro Ser Ala Glu Pro Gln Val His Arg Asn Ile His Ser
280 285 290
55 His Gly Leu Arg Ser Asn Leu Gln Glu Ile Phe Leu Pro Ala Phe Pro
295 300 305
Cys His Glu Arg Lys Lys Gln Glu
310 315

<210> 278
60 <211> 541
<212> PRT
<213> Homo sapiens

<220>

<221> SIGNAL

<222> -28..-1

5 <400> 278

	Met	Gly	Ser	Gln	Glu	Val	Leu	Gly	His	Ala	Ala	Arg	Leu	Ser	Ser	Ser
				-25					-20					-15		
	Gly	Leu	Leu	Leu	Gln	Val	Leu	Phe	Arg	Leu	Ile	Thr	Phe	Val	Leu	Asn
			-10					-5					1			
10	Ala	Phe	Ile	Leu	Arg	Phe	Leu	Ser	Lys	Glu	Ile	Val	Gly	Val	Val	Asn
	5				10					15						20
	Val	Arg	Leu	Thr	Leu	Leu	Tyr	Ser	Thr	Thr	Leu	Phe	Leu	Ala	Arg	Glu
				25						30					35	
15	Ala	Phe	Arg	Arg	Ala	Cys	Leu	Ser	Gly	Gly	Thr	Gln	Arg	Asp	Trp	Ser
				40					45					50		
	Gln	Thr	Leu	Asn	Leu	Leu	Trp	Leu	Thr	Val	Pro	Leu	Gly	Val	Phe	Trp
			55					60					65			
	Ser	Leu	Phe	Leu	Gly	Trp	Ile	Trp	Leu	Gln	Leu	Leu	Glu	Val	Pro	Asp
	70						75					80				
20	Pro	Asn	Val	Val	Pro	His	Tyr	Ala	Thr	Gly	Val	Val	Leu	Phe	Gly	Leu
	85					90					95					100
	Ser	Ala	Val	Val	Glu	Leu	Leu	Gly	Glu	Pro	Phe	Trp	Val	Leu	Ala	Gln
					105					110					115	
25	Ala	His	Met	Phe	Val	Lys	Leu	Lys	Val	Ile	Ala	Glu	Ser	Leu	Ser	Val
				120					125					130		
	Ile	Leu	Lys	Thr	Val	Leu	Thr	Ala	Phe	Leu	Val	Leu	Trp	Leu	Pro	His
			135					140					145			
	Trp	Gly	Leu	Tyr	Ile	Phe	Ser	Leu	Ala	Gln	Leu	Phe	Tyr	Thr	Thr	Val
		150					155					160				
30	Leu	Val	Leu	Cys	Tyr	Val	Ile	Tyr	Phe	Thr	Lys	Leu	Leu	Gly	Ser	Pro
	165					170					175					180
	Glu	Ser	Thr	Lys	Leu	Gln	Thr	Leu	Pro	Val	Ser	Arg	Ile	Thr	Asp	Leu
				185						190					195	
35	Leu	Pro	Asn	Ile	Thr	Arg	Asn	Gly	Ala	Phe	Ile	Asn	Trp	Lys	Glu	Ala
			200						205					210		
	Lys	Leu	Thr	Trp	Ser	Phe	Phe	Lys	Gln	Ser	Phe	Leu	Lys	Gln	Ile	Leu
			215					220					225			
	Thr	Glu	Gly	Glu	Arg	Tyr	Val	Met	Thr	Phe	Leu	Asn	Val	Leu	Asn	Phe
		230					235					240				
40	Gly	Asp	Gln	Gly	Val	Tyr	Asp	Ile	Val	Asn	Asn	Leu	Gly	Ser	Leu	Val
	245					250					255					260
	Ala	Arg	Leu	Ile	Phe	Gln	Pro	Ile	Glu	Glu	Ser	Phe	Tyr	Ile	Phe	Phe
				265						270					275	
45	Ala	Lys	Val	Leu	Glu	Arg	Gly	Lys	Asp	Ala	Thr	Leu	Gln	Lys	Gln	Glu
			280						285					290		
	Asp	Val	Ala	Val	Ala	Ala	Ala	Val	Leu	Glu	Ser	Leu	Leu	Lys	Leu	Ala
			295					300					305			
	Leu	Leu														

Leu His Leu Ser Pro Val Leu Leu Gly Thr Phe Ala Leu Ser Gly Gly
 440 445 450
 Val Thr Ala Val Ser Glu Val Phe Leu Cys Cys Glu Gln Gly Trp Pro
 455 460 465
 5 Ala Arg Leu Ala His Ile Ala Val Gly Ala Phe Cys Leu Gly Ala Thr
 470 475 480
 Leu Gly Thr Ala Phe Leu Thr Glu Thr Lys Leu Ile His Phe Leu Arg
 485 490 495 500
 Thr Gln Leu Gly Val Pro Arg Arg Thr Asp Lys Met Thr
 10 505 510

<210> 279

<211> 267

<212> PRT

15 <213> Homo sapiens

<220>

<221> SIGNAL

<222> -24...-1

20

<400> 279

Met Ala Arg Phe Leu Thr Leu Cys Thr Trp Leu Leu Leu Leu Gly Pro
 -20 -15 -10
 Gly Leu Leu Ala Thr Val Arg Ala Glu Cys Ser Gln Asp Cys Ala Thr
 -5 1 5
 25 Cys Ser Tyr Arg Leu Val Arg Pro Ala Asp Ile Asn Phe Leu Ala Cys
 10 15 20
 Val Met Glu Cys Glu Gly Lys Leu Pro Ser Leu Lys Ile Trp Glu Thr
 25 30 35 40
 30 Cys Lys Glu Leu Leu Gln Leu Ser Lys Pro Asp Leu Pro Gln Asp Gly
 45 50 55
 Thr Ser Thr Leu Arg Glu Asn Ser Lys Pro Glu Glu Ser His Leu Leu
 60 65 70
 Ala Lys Arg Tyr Gly Gly Phe Met Lys Arg Tyr Gly Gly Phe Met Lys
 75 80 85
 35 Lys Met Asp Glu Leu Tyr Pro Met Glu Pro Glu Glu Ala Asn Gly
 90 95 100
 Ser Glu Ile Leu Ala Lys Arg Tyr Gly Gly Phe Met Lys Lys Asp Ala
 105 110 115 120
 40 Glu Glu Asp Asp Ser Leu Ala Asn Ser Ser Asp Leu Leu Lys Glu Leu
 125 130 135
 Leu Glu Thr Gly Asp Asn Arg Glu Arg Ser His His Gln Asp Gly Ser
 140 145 150
 Asp Asn Glu Glu Val Ser Lys Arg Tyr Gly Gly Phe Met Arg Gly
 155 160 165
 45 Leu Lys Arg Ser Pro Gln Leu Glu Asp Glu Ala Lys Glu Leu Gln Lys
 170 175 180
 Arg Tyr Gly Gly Phe Met Arg Arg Val Gly Arg Pro Glu Trp Trp Met
 185 190 195 200
 50 Asp Tyr Gln Lys Arg Tyr Gly Gly Phe Leu Lys Arg Phe Ala Glu Ala
 205 210 215
 Leu Pro Ser Asp Glu Glu Gly Glu Ser Tyr Ser Lys Glu Val Pro Glu
 220 225 230
 Met Glu Lys Arg Tyr Gly Gly Phe Met Arg Phe
 55 235 240

<210> 280

<211> 362

<212> PRT

60 <213> Homo sapiens

<220>

<221> SIGNAL

<222> -40...-1

<400> 280

5 Met Pro Phe Ala Tyr Phe Phe Thr Glu Ser Glu Gly Phe Ala Gly Ser
 -40 -35 -30 -25
 Arg Lys Gly Val Leu Gly Arg Val Tyr Glu Thr Val Val Met Leu Met
 -20 -15 -10
 Leu Leu Thr Leu Leu Val Leu Gly Met Val Trp Val Ala Ser Ala Ile
 -5 1 5
 10 Val Asp Lys Asn Lys Ala Asn Arg Glu Ser Leu Tyr Asp Phe Trp Glu
 10 15 20
 Tyr Tyr Leu Pro Tyr Leu Tyr Ser Cys Ile Ser Phe Leu Gly Val Leu
 25 30 35 40
 Leu Leu Leu Val Cys Thr Pro Leu Gly Leu Ala Arg Met Phe Ser Val
 15 45 50 55
 Thr Gly Lys Leu Leu Val Lys Pro Arg Leu Leu Glu Asp Leu Glu Glu
 60 65 70
 Gln Leu Tyr Cys Ser Ala Phe Glu Glu Ala Ala Leu Thr Arg Arg Ile
 75 80 85
 20 Cys Asn Pro Thr Ser Cys Trp Leu Pro Leu Asp Met Glu Leu Leu His
 90 95 100
 Arg Gln Val Leu Ala Leu Gln Thr Gln Arg Val Leu Leu Glu Lys Arg
 105 110 115 120
 Arg Lys Ala Ser Ala Trp Gln Arg Asn Leu Gly Tyr Pro Leu Ala Met
 25 125 130 135
 Leu Cys Leu Leu Val Leu Thr Gly Leu Ser Val Leu Ile Val Ala Ile
 140 145 150
 His Ile Leu Glu Leu Leu Ile Asp Glu Ala Ala Met Pro Arg Gly Met
 155 160 165
 30 Gln Gly Thr Ser Leu Gly Gln Val Ser Phe Ser Lys Leu Gly Ser Phe
 170 175 180
 Gly Ala Val Ile Gln Val Val Leu Ile Phe Tyr Leu Met Val Ser Ser
 185 190 195 200
 Val Val Gly Phe Tyr Ser Ser Pro Leu Phe Arg Ser Leu Arg Pro Arg
 35 205 210 215
 Trp His Asp Thr Ala Met Thr Gln Ile Ile Gly Asn Cys Val Cys Leu
 220 225 230
 Leu Val Leu Ser Ser Ala Leu Pro Val Phe Ser Arg Thr Leu Gly Leu
 235 240 245
 40 Thr Arg Phe Asp Leu Leu Gly Asp Phe Gly Arg Phe Asn Trp Leu Gly
 250 255 260
 Asn Phe Tyr Ile Val Phe Leu Tyr Asn Ala Ala Phe Ala Gly Leu Thr
 265 270 275 280
 Thr Leu Tyr Leu Val Lys Thr Phe Thr Ala Ala Val Arg Ala Glu Leu
 45 285 290 295
 Ile Arg Ala Phe Gly Leu Asp Arg Leu Pro Leu Pro Val Ser Gly Phe
 300 305 310
 Pro Gln Ala Ser Arg Lys Thr Gln His Gln
 315 320

50

<210> 281

<211> 81

<212> PRT

<213> Homo sapiens

55

<220>

<221> SIGNAL

<222> -21...-1

60

<400> 281

Met Ser Arg Ser Ser Lys Val Val Leu Gly Leu Ser Val Leu Leu Thr
 -20 -15 -10
 Ala Ala Thr Val Ala Gly Val His Val Lys Gln Gln Trp Asp Gln Gln

```

-5          1          5          10
Arg Leu Arg Asp Gly Val Ile Arg Asp Ile Glu Arg Gln Ile Arg Lys
15          20          25
Lys Glu Asn Ile Arg Leu Leu Gly Glu Gln Ile Ile Leu Thr Glu Gln
5          30          35          40
Leu Glu Ala Glu Arg Glu Lys Met Leu Leu Ala Lys Gly Ser Gln Lys
45          50          55
Ser
60
10
<210> 282
<211> 541
<212> PRT
<213> Homo sapiens
15
<220>
<221> SIGNAL
<222> -28...-1
20
<400> 282
Met Gly Ser Gln Glu Val Leu Gly His Ala Ala Arg Leu Ala Ser Ser
-25 -20 -15
Gly Leu Leu Leu Gln Val Leu Phe Arg Leu Ile Thr Phe Val Leu Asn
-10 -5 1
25 Ala Phe Ile Leu Arg Phe Leu Ser Lys Glu Ile Val Gly Val Val Asn
5 10 15 20
Val Arg Leu Thr Leu Leu Tyr Ser Thr Thr Leu Phe Leu Ala Arg Glu
25 30 35
Ala Phe Arg Arg Ala Cys Leu Ser Gly Thr Gln Arg Asp Trp Ser
30 40 45 50
Gln Thr Leu Asn Leu Leu Trp Leu Thr Val Pro Leu Gly Val Phe Trp
55 60 65
Ser Leu Phe Leu Gly Trp Ile Trp Leu Gln Leu Leu Glu Val Pro Asp
70 75 80
35 Pro Asn Val Val Pro His Tyr Ala Thr Gly Val Val Leu Phe Gly Leu
85 90 95 100
Ser Ala Val Val Glu Leu Leu Gly Glu Pro Phe Trp Val Leu Ala Gln
105 110 115
Ala His Met Phe Val Lys Leu Lys Val Ile Ala Glu Ser Leu Ser Val
120 125 130
40 Ile Leu Lys Ser Val Leu Thr Ala Phe Leu Val Leu Trp Leu Pro His
135 140 145
Trp Gly Leu Tyr Ile Phe Ser Leu Ala Gln Leu Phe Tyr Thr Thr Val
150 155 160
45 Leu Val Leu Cys Tyr Val Ile Tyr Phe Thr Lys Leu Leu Gly Ser Pro
165 170 175 180
Glu Ser Thr Lys Leu Gln Thr Leu Pro Val Ser Arg Ile Thr Asp Leu
185 190 195
Leu Pro Asn Ile Thr Arg Asn Gly Ala Phe Ile Asn Trp Lys Glu Ala
200 205 210
50 Lys Leu Thr Trp Ser Phe Phe Lys Gln Ser Phe Leu Lys Gln Ile Leu
215 220 225
Thr Glu Gly Glu Arg Tyr Val Met Thr Phe Leu Asn Val Leu Asn Phe
230 235 240
55 Gly Asp Gln Gly Val Tyr Asp Ile Val Asn Asn Leu Gly Ser Leu Val
245 250 255 260
Ala Arg Leu Ile Phe Gln Pro Ile Glu Glu Ser Phe Tyr Ile Phe Phe
265 270 275
Ala Lys Val Leu Glu Arg Gly Lys Asp Ala Thr Leu Gln Lys Gln Glu
280 285 290
60 Asp Val Ala Val Ala Ala Ala Val Leu Glu Ser Leu Leu Lys Leu Ala
295 300 305
Leu Leu Ala Gly Leu Thr Ile Thr Val Phe Gly Phe Ala Tyr Ser Gln

```

310 315 320
 Leu Ala Leu Asp Ile Asn Gly Gly Thr Met Leu Ser Ser Gly Ser Gly
 325 330 335 340
 Pro Val Leu Leu Arg Ser Tyr Cys Leu Tyr Val Leu Leu Leu Ala Ile
 5 345 350 355
 Asn Gly Val Thr Glu Cys Phe Thr Phe Ala Ala Met Ser Lys Glu Glu
 360 365 370
 Val Asp Arg Tyr Asn Phe Val Met Leu Ala Leu Ser Ser Ser Phe Leu
 375 380 385
 10 Val Leu Ser Tyr Leu Leu Thr Arg Trp Cys Gly Ser Val Gly Phe Ile
 390 395 400
 Leu Ala Asn Cys Phe Asn Met Gly Ile Arg Ile Thr Gln Ser Leu Cys
 405 410 415 420
 Phe Ile His Arg Tyr Tyr Arg Arg Ser Pro His Arg Pro Leu Ala Gly
 15 425 430 435
 Leu His Leu Ser Pro Val Leu Leu Gly Thr Phe Ala Leu Ser Gly Gly
 440 445 450
 Val Thr Ala Val Ser Glu Val Phe Leu Cys Cys Glu Gln Gly Trp Pro
 455 460 465
 20 Ala Arg Leu Ala His Ile Ala Val Gly Ala Phe Cys Leu Gly Ala Thr
 470 475 480
 Leu Gly Thr Ala Phe Leu Thr Glu Thr Lys Leu Ile His Phe Leu Arg
 485 490 495 500
 Thr Gln Leu Gly Val Pro Arg Arg Thr Asp Lys Met Thr
 25 505 510

 <210> 283
 <211> 468
 <212> PRT
 30 <213> Homo sapiens

 <220>
 <221> SIGNAL
 <222> -21...-1
 35
 <400> 283
 Met Gly Thr Gln Glu Gly Trp Cys Leu Leu Leu Cys Leu Ala Leu Ser
 -20 -15 -10
 Gly Ala Ala Glu Thr Lys Pro His Pro Ala Glu Gly Gln Trp Arg Ala
 40 -5 1 5 10
 Val Asp Val Val Leu Asp Cys Phe Leu Val Lys Asp Gly Ala His Arg
 15 20 25
 Gly Ala Leu Ala Ser Ser Glu Asp Arg Ala Arg Ala Ser Leu Val Leu
 30 35 40
 45 Lys Gln Val Pro Val Leu Asp Asp Gly Ser Leu Glu Asp Phe Thr Asp
 45 50 55
 Phe Gln Gly Gly Thr Leu Ala Gln Asp Asp Pro Pro Ile Ile Phe Glu
 60 65 70 75
 Ala Ser Val Asp Leu Val Gln Ile Pro Gln Ala Glu Ala Leu Leu His
 50 80 85 90
 Ala Asp Cys Ser Gly Lys Glu Val Thr Cys Glu Ile Ser Arg Tyr Phe
 95 100 105
 Leu Gln Met Thr Glu Thr Thr Val Lys Thr Ala Ala Trp Phe Met Ala
 110 115 120
 55 Asn Val Gln Val Ser Gly Gly Pro Ser Ile Ser Leu Val Met Lys
 125 130 135
 Thr Pro Arg Val Ala Lys Asn Glu Val Leu Trp His Pro Thr Leu Asn
 140 145 150 155
 Leu Pro Leu Ser Pro Gln Gly Thr Val Arg Thr Ala Val Glu Phe Gln
 60 160 165 170
 Val Met Thr Gln Thr Gln Ser Leu Ser Phe Leu Leu Gly Ser Ser Ala
 175 180 185
 Ser Leu Asp Cys Gly Phe Ser Met Ala Pro Gly Leu Asp Leu Ile Ser

190 195 200
 Val Glu Trp Arg Leu Gln His Lys Gly Arg Gly Gln Leu Val Tyr Ser
 205 210 215
 Trp Thr Ala Gly Gln Gly Gln Ala Val Arg Lys Gly Ala Thr Leu Glu
 5 220 225 230 235
 Pro Ala Gln Leu Gly Met Ala Arg Asp Ala Ser Leu Thr Leu Pro Gly
 240 245 250
 Leu Thr Ile Gln Asp Glu Gly Thr Tyr Ile Cys Gln Ile Thr Thr Ser
 255 260 265
 10 Leu Tyr Arg Ala Gln Gln Ile Ile Gln Leu Asn Ile Gln Ala Ser Pro
 270 275 280
 Lys Val Arg Leu Ser Leu Ala Asn Glu Ala Leu Leu Pro Thr Leu Ile
 285 290 295
 Cys Asp Ile Ala Gly Tyr Tyr Pro Leu Asp Val Val Val Thr Trp Thr
 15 300 305 310 315
 Arg Glu Glu Leu Gly Gly Ser Pro Ala Gln Val Ser Gly Ala Ser Phe
 320 325 330
 Ser Ser Leu Arg Gln Ser Val Ala Gly Thr Tyr Ser Ile Ser Ser Ser
 335 340 345
 20 Leu Thr Ala Glu Pro Gly Ser Ala Gly Ala Thr Tyr Thr Cys Gln Val
 350 355 360
 Thr His Ile Ser Leu Glu Glu Pro Leu Gly Ala Ser Thr Gln Val Val
 365 370 375
 Pro Pro Glu Arg Arg Thr Ala Leu Gly Val Ile Phe Ala Ser Ser Leu
 25 380 385 390 395
 Phe Leu Leu Ala Leu Met Phe Leu Gly Leu Gln Arg Arg Gln Ala Pro
 400 405 410
 Thr Gly Leu Gly Leu Leu Gln Ala Glu Arg Trp Glu Thr Thr Ser Cys
 415 420 425
 30 Ala Asp Thr Gln Ser Ser His Leu His Glu Asp Arg Thr Ala Arg Val
 430 435 440
 Ser Gln Pro Ser
 445
 35 <210> 284
 <211> 406
 <212> PRT
 <213> Homo sapiens
 40 <220>
 <221> SIGNAL
 <222> -31...-1
 <400> 284
 45 Met Val Arg Ile Gln Arg Arg Lys Leu Leu Ala Ser Cys Leu Cys Val
 -30 -25 -20
 Thr Ala Thr Val Phe Leu Leu Val Thr Leu Gln Ala Leu Asp Thr Val
 -15 -10 -5 1
 Glu Asn Leu Met Lys Val Thr Gly Pro Pro Gln Gly Val Thr Asp Ser
 5 10 15
 50 Met Gln Cys Phe Asn Asp Gln Trp Pro Leu Ser Asn Thr Arg Ser Ser
 20 25 30
 Glu His Ile Lys Glu Val Met Val Glu Leu Gly Lys Phe Glu Arg Lys
 35 40 45
 55 Glu Phe Lys Ser Ser Ser Leu Gln Asp Gly His Thr Lys Met Glu Glu
 50 55 60 65
 Ala Pro Thr His Leu Asn Ser Phe Leu Lys Lys Glu Gly Leu Thr Phe
 70 75 80
 Asn Arg Lys Arg Lys Trp Glu Leu Asp Ser Tyr Pro Ile Met Leu Trp
 85 90 95
 60 Trp Ser Pro Leu Thr Gly Glu Thr Gly Arg Leu Gly Gln Cys Gly Ala
 100 105 110
 Asp Ala Cys Phe Phe Thr Ile Asn Arg Thr Tyr Leu His His His Met

115 120 125
 Thr Lys Ala Phe Leu Phe Tyr Gly Thr Asp Phe Asn Ile Asp Ser Leu
 130 135 140 145
 Pro Leu Pro Arg Lys Ala His His Asp Trp Ala Val Phe His Glu Glu
 150 155 160
 5 Ser Pro Lys Asn Asn Tyr Lys Leu Phe His Lys Pro Val Ile Thr Leu
 165 170 175
 Phe Asn Tyr Thr Ala Thr Phe Ser Arg His Ser His Leu Pro Leu Thr
 180 185 190
 10 Thr Gln Tyr Leu Glu Ser Ile Glu Val Leu Lys Ser Leu Arg Tyr Leu
 195 200 205
 Val Pro Leu Gln Ser Lys Asn Lys Leu Arg Lys Arg Leu Ala Pro Leu
 210 215 220 225
 Val Tyr Val Gln Ser Tyr Cys Asp Pro Pro Ser Asp Arg Asp Ser Tyr
 230 235 240
 15 Val Arg Glu Leu Met Thr Tyr Ile Glu Val Asp Ser Tyr Gly Glu Cys
 245 250 255
 Leu Arg Asn Lys Asp Leu Pro Gln Gln Leu Lys Asn Pro Ala Ser Met
 260 265 270
 20 Asp Ala Asp Gly Phe Tyr Arg Ile Ile Ala Gln Tyr Lys Phe Ile Leu
 275 280 285
 Ala Phe Glu Asn Ala Val Cys Asp Asp Tyr Ile Thr Glu Lys Phe Trp
 290 295 300 305
 Arg Pro Leu Lys Leu Gly Val Val Pro Val Tyr Tyr Gly Ser Pro Ser
 310 315 320
 25 Ile Thr Asp Trp Leu Pro Ser Asn Lys Ser Ala Ile Leu Val Ser Glu
 325 330 335
 Phe Ser His Pro Arg Glu Leu Ala Ser Tyr Ile Arg Arg Leu Asp Ser
 340 345 350
 30 Asp Asp Arg Leu Tyr Glu Ala Tyr Val Glu Trp Lys Leu Lys Gly Arg
 355 360 365
 Ser Leu Thr Ser Asp Phe
 370 375

 35 <210> 285
 <211> 305
 <212> PRT
 <213> Homo sapiens

 40 <220>
 <221> SIGNAL
 <222> -26...-1

 <400> 285
 45 Met Gly Ile Gln Thr Ser Pro Val Leu Leu Ala Ser Leu Gly Val Gly
 -25 -20 -15
 Leu Val Thr Leu Leu Gly Leu Ala Val Gly Ser Tyr Leu Val Arg Arg
 -10 -5 1 5
 Ser Arg Arg Pro Gln Val Thr Leu Leu Asp Pro Asn Glu Lys Tyr Leu
 10 15 20
 50 Leu Arg Leu Leu Asp Lys Thr Thr Val Ser His Asn Thr Lys Arg Phe
 25 30 35
 Arg Phe Ala Leu Pro Thr Ala His His Thr Leu Gly Leu Pro Val Gly
 40 45 50
 55 Lys His Ile Tyr Leu Ser Thr Arg Ile Asp Gly Ser Leu Val Ile Arg
 55 60 65 70
 Pro Tyr Thr Pro Val Thr Ser Asp Glu Asp Gln Gly Tyr Val Asp Leu
 75 80 85
 Val Ile Lys Val Tyr Leu Lys Gly Val His Pro Lys Phe Pro Glu Gly
 90 95 100
 60 Gly Lys Met Ser Gln Tyr Leu Asp Ser Leu Lys Val Gly Asp Val Val
 105 110 115
 Glu Phe Arg Gly Pro Ser Gly Leu Leu Thr Tyr Thr Gly Lys Gly His

120 125 130
 Phe Asn Ile Gln Pro Asn Lys Lys Ser Pro Pro Glu Pro Arg Val Ala
 135 140 145 150
 Lys Lys Leu Gly Met Ile Ala Gly Gly Thr Gly Ile Thr Pro Met Leu
 5 155 160 165
 Gln Leu Ile Arg Ala Ile Leu Lys Val Pro Glu Asp Pro Thr Gln Cys
 170 175 180
 Phe Leu Leu Phe Ala Asn Gln Thr Glu Lys Asp Ile Ile Leu Arg Glu
 185 190 195
 10 Asp Leu Glu Glu Leu Gln Ala Arg Tyr Pro Asn Arg Phe Lys Leu Trp
 200 205 210
 Phe Thr Leu Asp His Pro Pro Lys Asp Trp Ala Tyr Ser Lys Gly Phe
 215 220 225 230
 Val Thr Ala Asp Met Ile Arg Glu His Leu Pro Ala Pro Gly Asp Asp
 15 235 240 245
 Val Leu Val Leu Leu Cys Gly Pro Pro Pro Met Val Gln Leu Ala Cys
 250 255 260
 His Pro Asn Leu Asp Lys Leu Gly Tyr Ser Gln Lys Met Arg Phe Thr
 265 270 275
 20 Tyr

 <210> 286
 <211> 442
 <212> PRT
 25 <213> Homo sapiens

 <220>
 <221> SIGNAL
 <222> -21...-1
 30
 <220>
 <221> UNSURE
 <222> 132
 <223> Xaa = Pro, Arg
 35
 <400> 286
 Met Gly Thr Gln Glu Gly Trp Cys Leu Leu Leu Cys Leu Ala Leu Ser
 -20 -15 -10
 Gly Ala Ala Glu Thr Lys Pro His Pro Ala Glu Gly Gln Leu Arg Ala
 40 -5 1 5 10
 Val Asp Val Val Leu Asp Cys Phe Leu Ala Lys Asp Gly Ala His Arg
 15 20 25
 Gly Ala Leu Ala Ser Ser Glu Asp Arg Ala Arg Ala Ser Leu Val Leu
 30 35 40
 45 Lys Gln Val Pro Val Leu Asp Asp Gly Ser Leu Glu Asp Phe Thr Asp
 45 50 55
 Phe Gln Gly Gly Thr Leu Ala Gln Asp Asp Pro Pro Ile Ile Phe Glu
 60 65 70 75
 Ala Ser Val Asp Leu Val Gln Ile Pro Gln Ala Glu Ala Leu Leu His
 50 80 85 90
 Ala Asp Cys Ser Gly Lys Glu Val Thr Cys Glu Ile Ser Arg Tyr Phe
 95 100 105
 Leu Gln Met Thr Glu Thr Thr Val Lys Thr Ala Ala Trp Phe Met Ala
 110 115 120
 55 Asn Met Gln Val Ser Gly Gly Xaa Ser Ile Ser Leu Val Met Lys
 125 130 135
 Thr Pro Arg Val Thr Lys Asn Glu Ala Leu Trp His Pro Thr Leu Asn
 140 145 150 155
 Leu Pro Leu Ser Pro Gln Gly Thr Val Arg Thr Ala Val Glu Phe Gln
 60 160 165 170
 Val Met Thr Gln Thr Gln Ser Leu Ser Phe Leu Leu Gly Ser Ser Ala
 175 180 185
 Ser Leu Asp Cys Gly Phe Ser Met Ala Pro Gly Leu Asp Leu Ile Ser

		190				195				200				
	Val	Glu	Trp	Arg	Leu	Gln	His	Lys	Gly	Arg	Gly	Gln	Leu	Val
		205					210					215		
	Trp	Thr	Ala	Gly	Gln	Gly	Gln	Ala	Val	Arg	Lys	Gly	Ala	Thr
5	220					225					230			235
	Pro	Ala	Gln	Leu	Gly	Met	Ala	Arg	Asp	Ala	Ser	Leu	Thr	Leu
					240					245				250
	Leu	Thr	Ile	Gln	Asp	Glu	Gly	Thr	Tyr	Ile	Cys	Gln	Ile	Thr
				255					260					265
10	Leu	Tyr	Arg	Ala	Gln	Gln	Ile	Ile	Gln	Leu	Asn	Ile	Gln	Ala
			270					275					280	
	Lys	Val	Arg	Leu	Ser	Leu	Ala	Asn	Glu	Ala	Leu	Leu	Pro	Thr
		285					290					295		
	Cys	Asp	Ile	Ala	Gly	Tyr	Tyr	Pro	Leu	Asp	Val	Val	Val	Thr
15	300					305					310			315
	Arg	Glu	Glu	Leu	Gly	Gly	Ser	Pro	Ala	Gln	Val	Ser	Gly	Ala
					320					325				330
	Ser	Ser	Leu	Arg	Gln	Ser	Val	Ala	Gly	Thr	Tyr	Ser	Ile	Ser
				335					340					345
20	Leu	Thr	Ala	Glu	Pro	Gly	Ser	Ala	Gly	Ala	Thr	Tyr	Thr	Cys
			350					355						360
	Thr	His	Ile	Ser	Leu	Glu	Glu	Pro	Leu	Gly	Ala	Ser	Thr	Gln
		365					370					375		
	Pro	Pro	Glu	Arg	Arg	Thr	Ala	Leu	Gly	Val	Ile	Phe	Ala	Ser
25	380					385					390			395
	Phe	Leu	Leu	Ala	Leu	Met	Phe	Leu	Gly	Leu	Gln	Arg	Arg	Gln
				400						405				410
	Thr	Gly	Leu	Gly	Leu	Leu	Gln	Ala	Glu	Arg				
				415					420					
30														
	<210>	287												
	<211>	286												
	<212>	PRT												
	<213>	Homo sapiens												
35														
	<220>													
	<221>	SIGNAL												
	<222>	-48...-1												
40	<400>	287												
	Met	Asn	Pro	Ala	Ser	Asp	Gly	Gly	Thr	Ser	Glu	Ser	Ile	Phe
				-45					-40				-35	
	Asp	Tyr	Ala	Ser	Trp	Gly	Ile	Arg	Ser	Thr	Leu	Met	Val	Ala
			-30					-25					-20	
45	Val	Phe	Tyr	Leu	Gly	Val	Phe	Val	Val	Cys	His	Gln	Leu	Ser
		-15				-10					-5			
	Leu	Asn	Ala	Thr	Tyr	Arg	Ser	Leu	Val	Ala	Arg	Glu	Lys	Val
	1			5					10				15	
	Asp	Leu	Ala	Ala	Thr	Arg	Ala	Val	Phe	Gly	Val	Gln	Ser	Thr
50			20					25					30	
	Gly	Leu	Trp	Ala	Leu	Leu	Gly	Asp	Pro	Val	Leu	His	Ala	Asp
		35					40					45		
	Arg	Gly	Gln	Gln	Asn	Trp	Cys	Trp	Phe	His	Ile	Thr	Ala	Thr
		50				55					60			
55	Phe	Phe	Cys	Phe	Glu	Asn	Val	Ala	Val	His	Leu	Ser	Asn	Leu
	65				70					75				80
	Arg	Thr	Phe	Asp	Leu	Phe	Leu	Val	Ile	His	His	Leu	Phe	Ala
			85					90					95	
	Gly	Phe	Leu	Gly	Cys	Leu	Val	Asn	Leu	Gln	Ala	Gly	His	Tyr
60			100					105					110	
	Met	Thr	Thr	Leu	Leu	Leu	Glu	Met	Ser	Thr	Pro	Phe	Thr	Cys
		115						120					125	
	Trp	Met	Leu	Leu	Lys	Ala	Gly	Trp	Ser	Glu	Ser	Leu	Phe	Trp

130 135 140
 Asn Gln Trp Leu Met Ile His Met Phe His Cys Arg Met Val Leu Thr
 145 150 155 160
 Tyr His Met Trp Trp Val Cys Phe Trp His Trp Asp Gly Leu Val Ser
 5 165 170 175
 Ser Leu Tyr Leu Pro His Leu Thr Leu Phe Leu Val Gly Leu Ala Leu
 180 185 190
 Leu Thr Leu Ile Ile Asn Pro Tyr Trp Thr His Lys Lys Thr Gln Gln
 195 200 205
 10 Leu Leu Asn Pro Val Asp Trp Asn Phe Ala Gln Pro Glu Ala Lys Ser
 210 215 220
 Arg Pro Glu Gly Asn Gly Gln Leu Leu Arg Lys Lys Arg Pro
 225 230 235

 15 <210> 288
 <211> 398
 <212> PRT
 <213> Homo sapiens

 20 <220>
 <221> SIGNAL
 <222> -21...-1

 <400> 288
 25 Met Val Asn Asp Pro Pro Val Pro Ala Leu Leu Trp Ala Gln Glu Val
 -20 -15 -10
 Gly Gln Val Leu Ala Gly Arg Ala Arg Arg Leu Leu Leu Gln Phe Gly
 -5 1 5 10
 Val Leu Phe Cys Thr Ile Leu Leu Leu Trp Val Ser Val Phe Leu
 30 15 20 25
 Tyr Gly Ser Phe Tyr Tyr Ser Tyr Met Pro Thr Val Ser His Leu Ser
 30 35 40
 Pro Val His Phe Tyr Tyr Arg Thr Asp Cys Asp Ser Ser Thr Thr Ser
 45 50 55
 35 Leu Cys Ser Phe Pro Val Ala Asn Val Ser Leu Thr Lys Gly Gly Arg
 60 65 70 75
 Asp Arg Val Leu Met Tyr Gly Gln Pro Tyr Arg Val Thr Leu Glu Leu
 80 85 90
 Glu Leu Pro Glu Ser Pro Val Asn Gln Asp Leu Gly Met Phe Leu Val
 40 95 100 105
 Thr Ile Ser Cys Tyr Thr Arg Gly Gly Arg Ile Ile Ser Thr Ser Ser
 110 115 120
 Arg Ser Val Met Leu His Tyr Arg Ser Asp Leu Leu Gln Met Leu Asp
 125 130 135
 45 Thr Leu Val Phe Ser Ser Leu Leu Leu Phe Gly Phe Ala Glu Gln Lys
 140 145 150 155
 Gln Leu Leu Glu Val Glu Leu Tyr Ala Asp Tyr Arg Glu Asn Ser Tyr
 160 165 170
 Val Pro Thr Thr Gly Ala Ile Ile Glu Ile His Ser Lys Arg Ile Gln
 50 175 180 185
 Leu Tyr Gly Ala Tyr Leu Arg Ile His Ala His Phe Thr Gly Leu Arg
 190 195 200
 Tyr Leu Leu Tyr Asn Phe Pro Met Thr Cys Ala Phe Ile Gly Val Ala
 205 210 215
 55 Ser Asn Phe Thr Phe Leu Ser Val Ile Val Leu Phe Ser Tyr Met Gln
 220 225 230 235
 Trp Val Trp Gly Gly Ile Trp Pro Arg His Arg Phe Ser Leu Gln Val
 240 245 250
 Asn Ile Arg Lys Arg Asp Asn Ser Arg Lys Glu Val Gln Arg Arg Ile
 255 260 265
 60 Ser Ala His Gln Pro Gly Pro Glu Gly Gln Glu Glu Ser Thr Pro Gln
 270 275 280
 Ser Asp Val Thr Glu Asp Gly Glu Ser Pro Glu Asp Pro Ser Gly Thr

```

      285              290              295
      Glu Gly Gln Leu Ser Glu Glu Lys Pro Asp Gln Gln Pro Leu Ser
      300              305              310              315
      Gly Glu Glu Glu Leu Glu Pro Glu Ala Ser Asp Gly Ser Gly Ser Trp
5      320              325              330
      Glu Asp Ala Ala Leu Leu Thr Glu Ala Asn Leu Pro Ala Pro Ala Pro
      335              340              345
      Ala Ser Ala Ser Ala Pro Val Leu Glu Thr Leu Gly Ser Ser Glu Pro
      350              355              360
10     Ala Gly Gly Ala Leu Arg Gln Arg Pro Thr Cys Ser Ser Ser
      365              370              375

```

```

      <210> 289
      <211> 130
15     <212> PRT
      <213> Homo sapiens

```

```

      <220>
      <221> SIGNAL
20     <222> -20...-1

```

```

      <400> 289
      Met Arg Gln Lys Ala Val Ser Leu Phe Phe Cys Tyr Leu Leu Leu Phe
      -20              -15              -10              -5
25     Thr Cys Ser Gly Val Glu Ala Gly Lys Lys Lys Cys Ser Glu Ser Ser
      1              5              10
      Asp Ser Gly Ser Gly Phe Trp Lys Ala Leu Thr Phe Met Ala Val Gly
      15              20              25
      Gly Gly Leu Ala Val Ala Gly Leu Pro Ala Leu Gly Phe Thr Gly Ala
30     30              35              40
      Gly Ile Ala Ala Asn Ser Val Ala Ala Ser Leu Met Ser Trp Ser Ala
      45              50              55              60
      Ile Leu Asn Gly Gly Gly Val Pro Ala Gly Gly Leu Val Ala Thr Leu
      65              70              75
35     Gln Ser Leu Gly Ala Gly Gly Ser Ser Val Val Ile Gly Asn Ile Gly
      80              85              90
      Ala Leu Met Gly Tyr Ala Thr His Lys Tyr Leu Asp Ser Glu Glu Asp
      95              100              105
      Glu Glu
40     110

```

```

      <210> 290
      <211> 86
      <212> PRT
45     <213> Homo sapiens

```

```

      <220>
      <221> SIGNAL
      <222> -20...-1

```

```

50     <400> 290
      Met Ala Val Gly Gly Gly Leu Ala Val Ala Gly Leu Pro Ala Leu Gly
      -20              -15              -10              -5
      Phe Thr Gly Ala Gly Ile Ala Ala Asn Ser Val Ala Ala Ser Leu Met
55     1              5              10
      Ser Trp Ser Ala Ile Leu Asn Gly Gly Gly Val Pro Ala Gly Gly Leu
      15              20              25
      Val Ala Thr Leu Gln Ser Leu Gly Ala Gly Gly Ser Ser Val Val Ile
      30              35              40
60     Gly Asn Ile Gly Ala Leu Met Gly Tyr Ala Thr His Lys Tyr Leu Asp
      45              50              55              60
      Ser Glu Glu Asp Glu Glu
      65

```

<210> 291
 <211> 207
 <212> PRT
 5 <213> Homo sapiens

 <220>
 <221> SIGNAL
 <222> -23...-1
 10
 <400> 291
 Met Ala Pro Phe Glu Pro Leu Ala Ser Gly Ile Leu Leu Leu Leu Trp
 -20 -15 -10
 Leu Ile Ala Pro Ser Arg Ala Cys Thr Cys Val Pro Pro His Pro Gln
 15 -5 1 5
 Thr Ala Phe Cys Asn Ser Asp Leu Val Ile Arg Ala Lys Phe Val Gly
 10 15 20 25
 Thr Pro Glu Val Asn Gln Thr Thr Leu Tyr Gln Arg Tyr Glu Ile Lys
 30 35 40
 20 Met Thr Lys Met Tyr Lys Gly Phe Gln Ala Leu Gly Asp Ala Ala Asp
 45 50 55
 Ile Arg Phe Val Tyr Thr Pro Ala Met Glu Ser Val Cys Gly Tyr Phe
 60 65 70
 His Arg Ser His Asn Arg Ser Glu Glu Phe Leu Ile Ala Gly Lys Leu
 25 75 80 85
 Gln Asp Gly Leu Leu His Ile Thr Thr Cys Ser Phe Val Ala Pro Trp
 90 95 100 105
 Asn Ser Leu Ser Leu Ala Gln Arg Arg Gly Phe Thr Lys Thr Tyr Thr
 110 115 120
 30 Val Gly Cys Glu Glu Cys Thr Val Phe Pro Cys Leu Ser Phe Pro Cys
 125 130 135
 Lys Leu Gln Ser Gly Thr His Cys Leu Trp Thr Asp Gln Leu Leu Gln
 140 145 150
 Gly Ser Glu Lys Gly Phe Gln Ser Arg His Leu Ala Cys Leu Pro Arg
 35 155 160 165
 Glu Pro Gly Leu Cys Thr Trp Gln Ser Leu Arg Ser Gln Ile Ala
 170 175 180

 <210> 292
 40 <211> 111
 <212> PRT
 <213> Homo sapiens

 <220>
 45 <221> SIGNAL
 <222> -24...-1

 <400> 292
 Met Lys Tyr Asp Cys Pro Phe Ser Gly Thr Ser Phe Val Val Phe Ser
 50 -20 -15 -10
 Leu Phe Leu Ile Cys Ala Met Ala Gly Asp Val Val Tyr Ala Asp Ile
 -5 1 5
 Lys Thr Val Arg Thr Ser Pro Leu Glu Leu Ala Phe Pro Leu Gln Arg
 10 15 20
 55 Ser Val Ser Phe Asn Phe Ser Thr Val His Lys Ser Cys Pro Ala Lys
 25 30 35 40
 Asp Trp Lys Val His Lys Gly Lys Cys Tyr Trp Ile Ala Glu Thr Lys
 45 50 55
 Lys Ser Trp Asn Lys Ser Gln Asn Asp Cys Ala Ile Asn Asn Ser Tyr
 60 60 65 70
 Leu Met Val Ile Gln Asp Ile Thr Ala Met Val Arg Phe Asn Ile
 75 80 85

<210> 293
 <211> 139
 <212> PRT
 <213> Homo sapiens

5

<220>
 <221> SIGNAL
 <222> -15...-1

10

<400> 293
 Met Glu Ala Val Val Phe Val Phe Ser Leu Leu Asp Cys Cys Ala Leu
 -15 -10 -5 1
 Ile Phe Leu Ser Val Tyr Phe Ile Ile Thr Leu Ser Asp Leu Glu Cys
 5 10 15
 15 Asp Tyr Ile Asn Ala Arg Ser Cys Cys Ser Lys Leu Asn Lys Trp Val
 20 25 30
 Ile Pro Glu Leu Ile Gly His Thr Ile Val Thr Val Leu Leu Leu Met
 35 40 45
 Ser Leu His Trp Phe Ile Phe Leu Leu Asn Leu Pro Val Ala Thr Trp
 20 50 55 60 65
 Asn Ile Tyr Arg Tyr Ile Met Val Pro Ser Gly Asn Met Gly Val Phe
 70 75 80
 Asp Pro Thr Glu Ile His Asn Arg Gly Gln Leu Lys Ser His Met Lys
 85 90 95
 25 Glu Ala Met Ile Lys Leu Gly Phe His Leu Leu Cys Phe Phe Met Tyr
 100 105 110
 Leu Tyr Ser Met Ile Leu Ala Leu Ile Asn Asp
 115 120

30

<210> 294
 <211> 160
 <212> PRT
 <213> Homo sapiens

35

<220>
 <221> SIGNAL
 <222> -27...-1

<400> 294
 40 Met Gln Arg Val Ser Gly Leu Leu Ser Trp Thr Leu Ser Arg Val Leu
 -25 -20 -15
 Trp Leu Ser Gly Leu Ser Glu Pro Gly Ala Ala Arg Gln Pro Arg Ile
 -10 -5 1 5
 Met Glu Glu Lys Ala Leu Glu Val Tyr Asp Leu Ile Arg Thr Ile Arg
 10 15 20
 45 Asp Pro Glu Lys Pro Asn Thr Leu Glu Glu Leu Glu Val Val Ser Glu
 25 30 35
 Ser Cys Val Glu Val Gln Glu Ile Asn Glu Glu Glu Tyr Leu Val Ile
 40 45 50
 50 Ile Arg Phe Thr Pro Thr Val Pro His Cys Ser Leu Ala Thr Leu Ile
 55 60 65
 Gly Leu Cys Leu Arg Val Lys Leu Gln Arg Cys Leu Pro Phe Lys His
 70 75 80 85
 Lys Leu Glu Ile Tyr Ile Ser Glu Gly Thr His Ser Thr Glu Glu Asp
 90 95 100
 55 Ile Asn Lys Gln Ile Asn Asp Lys Glu Arg Val Ala Ala Ala Met Glu
 105 110 115
 Asn Pro Asn Leu Arg Glu Ile Val Glu Gln Cys Val Leu Glu Pro Asp
 120 125 130

60

<210> 295
 <211> 181
 <212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

5 <222> -16...-1

<400> 295

```

Met Pro Pro Phe Leu Leu Leu Thr Cys Leu Phe Ile Thr Gly Thr Ser
-15 -10 -5
10 Val Ser Pro Val Ala Leu Asp Pro Cys Ser Ala Tyr Ile Ser Leu Asn
1 5 10 15
Glu Pro Trp Arg Asn Thr Asp His Gln Leu Asp Glu Ser Gln Gly Pro
20 25 30
Pro Leu Cys Asp Asn His Val Asn Gly Glu Trp Tyr His Phe Thr Gly
15 35 40 45
Met Ala Gly Asp Ala Met Pro Thr Phe Cys Ile Pro Glu Asn His Cys
50 55 60
Gly Thr His Ala Pro Val Trp Leu Asn Gly Ser His Pro Leu Glu Gly
65 70 75 80
20 Asp Gly Ile Val Gln Arg Gln Ala Cys Ala Ser Phe Asn Gly Asn Cys
85 90 95
Cys Leu Trp Asn Thr Thr Val Glu Val Lys Ala Cys Pro Gly Gly Tyr
100 105 110
Tyr Val Tyr Arg Leu Thr Lys Pro Ser Val Cys Phe His Val Tyr Cys
115 120 125
25 Gly Arg Glu Tyr Leu Pro Cys Ala Leu Phe Leu His Gln Gln Gly His
130 135 140
Arg Trp Ser Pro Lys Val Pro Asn Tyr Arg Ile Cys Ser Tyr Ser Gly
145 150 155 160
30 Asn Tyr Ile Ser Ile
165

```

<210> 296

<211> 247

35 <212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

40 <222> -18...-1

<400> 296

```

Met Gly Leu Pro Gly Leu Phe Cys Leu Ala Val Leu Ala Ala Ser Ser
-15 -10 -5
45 Phe Ser Lys Ala Arg Glu Glu Glu Ile Thr Pro Val Val Ser Ile Ala
1 5 10
Tyr Lys Val Leu Glu Val Phe Pro Lys Gly Arg Trp Val Leu Ile Thr
15 20 25 30
Cys Cys Ala Pro Gln Pro Pro Pro Pro Ile Thr Tyr Ser Leu Cys Gly
35 40 45
50 Thr Lys Asn Ile Lys Val Ala Lys Lys Val Val Lys Thr His Glu Pro
50 55 60
Ala Ser Phe Asn Leu Asn Val Thr Leu Lys Ser Ser Pro Asp Leu Leu
65 70 75
55 Thr Tyr Phe Cys Arg Ala Ser Ser Thr Ser Gly Ala His Val Asp Ser
80 85 90
Ala Arg Leu Gln Met His Trp Glu Leu Trp Ser Lys Pro Val Ser Glu
95 100 105 110
Leu Arg Ala Asn Phe Thr Leu Gln Asp Arg Gly Ala Gly Pro Arg Val
115 120 125
60 Glu Met Ile Cys Gln Ala Ser Ser Gly Ser Pro Pro Ile Thr Asn Ser
130 135 140
Leu Ile Gly Lys Asp Gly Gln Val His Leu Gln Gln Arg Pro Cys His

```

145 150 155
 Arg Gln Pro Ala Asn Phe Ser Phe Leu Pro Ser Gln Thr Ser Asp Trp
 160 165 170
 Phe Trp Cys Gln Ala Ala Asn Asn Ala Asn Val Gln His Ser Ala Leu
 5 175 180 185 190
 Thr Val Val Pro Pro Gly Gly Leu Pro Arg Ala Pro Thr Ile Val Leu
 195 200 205
 Val Gly Ser Leu Ala Ser Thr Ala Ala Ile Thr Ser Arg Met Leu Gly
 210 215 220
 10 Trp Thr Thr Trp Ala Arg Trp
 225

<210> 297

<211> 132

15 <212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

20 <222> -41...-1

<400> 297

Met Glu Gly Gly Ala Tyr Gly Ala Gly Lys Ala Gly Gly Ala Phe Asp
 -40 -35 -30
 25 Pro Tyr Thr Leu Val Arg Gln Pro His Thr Ile Leu Arg Val Val Ser
 -25 -20 -15 -10
 Trp Leu Phe Ser Ile Val Val Phe Gly Ser Ile Val Asn Glu Gly Tyr
 -5 1 5
 Leu Asn Ser Ala Ser Glu Gly Glu Gln Phe Cys Ile Tyr Asn Arg Asn
 10 15 20
 30 Pro Asn Ala Cys Ser Tyr Gly Val Ala Val Gly Val Leu Ala Phe Leu
 25 30 35
 Thr Cys Leu Leu Tyr Leu Ala Leu Asp Val Tyr Phe Pro Gln Ile Ser
 40 45 50 55
 35 Ser Val Lys Asp Arg Lys Lys Ala Val Leu Ser Asp Ile Gly Val Ser
 60 65 70
 Gly Glu Pro His Pro Ala Gly Thr Pro Cys Thr Glu Ser Thr Glu Gly
 75 80 85
 Cys Pro Gly Pro
 40 90

<210> 298

<211> 251

<212> PRT

45 <213> Homo sapiens

<220>

<221> SIGNAL

<222> -24...-1

50

<400> 298

Met Leu Gly Ala Arg Leu Arg Leu Trp Val Cys Ala Leu Cys Ser Val
 -20 -15 -10
 Cys Ser Met Ser Val Leu Arg Ala Tyr Pro Asn Ala Ser Pro Leu Leu
 -5 1 5
 55 Gly Ser Ser Trp Gly Gly Leu Ile His Leu Tyr Thr Ala Thr Ala Arg
 10 15 20
 Asn Ser Tyr His Leu Gln Ile His Lys Asn Gly His Val Asp Gly Ala
 25 30 35 40
 60 Pro His Gln Thr Ile Tyr Ser Ala Leu Met Ile Arg Ser Glu Asp Ala
 45 50 55
 Gly Phe Val Val Ile Thr Gly Val Met Ser Arg Arg Tyr Leu Cys Met
 60 65 70

Asp Phe Arg Gly Asn Ile Phe Gly Ser His Tyr Phe Asp Pro Glu Asn
 75 80 85
 Cys Arg Phe Gln His Gln Thr Leu Glu Asn Gly Tyr Asp Val Tyr His
 90 95 100
 5 Ser Pro Gln Tyr His Phe Leu Val Ser Leu Gly Arg Ala Lys Arg Ala
 105 110 115 120
 Phe Leu Pro Gly Met Asn Pro Pro Pro Tyr Ser Gln Phe Leu Ser Arg
 125 130 135
 10 Arg Asn Glu Ile Pro Leu Ile His Phe Asn Thr Pro Ile Pro Arg Arg
 140 145 150
 His Thr Arg Ser Ala Glu Asp Asp Ser Glu Arg Asp Pro Leu Asn Val
 155 160 165
 Leu Lys Pro Arg Ala Arg Met Thr Pro Ala Pro Ala Ser Cys Ser Gln
 170 175 180
 15 Glu Leu Pro Ser Ala Glu Asp Asn Ser Pro Met Ala Ser Asp Pro Leu
 185 190 195 200
 Gly Val Val Arg Gly Gly Arg Val Asn Thr His Ala Gly Gly Thr Gly
 205 210 215
 20 Pro Glu Gly Cys Arg Pro Phe Ala Lys Phe Ile
 220 225

 <210> 299
 <211> 137
 <212> PRT
 25 <213> Homo sapiens

 <220>
 <221> SIGNAL
 <222> -22...-1
 30
 <400> 299
 Met Leu Ser Gly Arg Leu Val Leu Gly Leu Val Ser Met Ala Gly Arg
 -20 -15 -10
 35 Val Cys Leu Cys Gln Gly Ser Ala Gly Ser Gly Ala Ile Gly Pro Val
 -5 1 5 10
 Glu Ala Ala Ile Arg Thr Lys Leu Glu Glu Ala Leu Ser Pro Glu Val
 15 20 25
 Leu Glu Leu Arg Asn Glu Ser Gly Gly His Ala Val Pro Pro Gly Ser
 30 35 40
 40 Glu Thr His Phe Arg Val Ala Val Val Ser Ser Arg Phe Glu Gly Leu
 45 50 55
 Ser Pro Leu Gln Arg His Arg Leu Val His Ala Ala Leu Ala Glu Glu
 60 65 70
 45 Leu Gly Gly Pro Val His Ala Leu Ala Ile Gln Ala Arg Thr Pro Ala
 75 80 85 90
 Gln Trp Arg Glu Asn Ser Gln Leu Asp Thr Ser Pro Pro Cys Leu Gly
 95 100 105
 Gly Asn Lys Lys Thr Leu Gly Thr Pro
 110 115
 50
 <210> 300
 <211> 541
 <212> PRT
 <213> Homo sapiens
 55
 <220>
 <221> SIGNAL
 <222> -28...-1
 60
 <400> 300
 Met Gly Ser Gln Glu Val Leu Gly His Ala Ala Arg Leu Ala Ser Ser
 -25 -20 -15
 Gly Leu Leu Leu Gln Val Leu Phe Arg Leu Ile Thr Phe Val Leu Asn

		-10				-5				1				
	Ala	Phe	Ile	Leu	Arg	Phe	Leu	Ser	Lys	Glu	Ile	Val	Gly	Val
	5					10					15			20
	Val	Arg	Leu	Thr	Leu	Leu	Tyr	Ser	Thr	Thr	Leu	Phe	Leu	Ala
5					25					30				35
	Ala	Phe	Arg	Arg	Ala	Cys	Leu	Ser	Gly	Gly	Thr	Gln	Arg	Asp
				40					45				50	
	Gln	Thr	Leu	Asn	Leu	Leu	Trp	Leu	Thr	Val	Pro	Leu	Gly	Val
			55					60				65		
10	Ser	Leu	Phe	Leu	Gly	Trp	Ile	Trp	Leu	Gln	Leu	Leu	Glu	Val
		70					75				80			
	Pro	Asn	Val	Val	Pro	His	Tyr	Ala	Thr	Gly	Val	Val	Leu	Phe
	85					90				95				100
	Ser	Ala	Val	Val	Glu	Leu	Leu	Gly	Glu	Pro	Phe	Trp	Val	Leu
15					105					110				115
	Ala	His	Met	Phe	Val	Lys	Leu	Lys	Val	Ile	Ala	Glu	Ser	Leu
				120					125					130
	Ile	Leu	Lys	Ser	Val	Leu	Thr	Ala	Phe	Leu	Val	Leu	Trp	Leu
			135					140				145		
20	Trp	Gly	Leu	Tyr	Ile	Phe	Ser	Leu	Ala	Gln	Leu	Phe	Tyr	Thr
		150					155				160			
	Leu	Val	Leu	Cys	Tyr	Val	Ile	Tyr	Phe	Thr	Lys	Leu	Leu	Gly
	165				170					175				180
	Glu	Ser	Thr	Lys	Leu	Gln	Thr	Leu	Pro	Val	Ser	Arg	Ile	Thr
25				185						190				195
	Leu	Pro	Asn	Ile	Thr	Arg	Asn	Gly	Ala	Phe	Ile	Asn	Trp	Lys
			200					205				210		
	Lys	Leu	Thr	Trp	Ser	Phe	Phe	Lys	Gln	Ser	Phe	Leu	Lys	Gln
		215					220					225		
30	Thr	Glu	Gly	Glu	Arg	Tyr	Val	Met	Thr	Phe	Leu	Asn	Val	Leu
		230					235					240		
	Gly	Asp	Gln	Gly	Val	Tyr	Asp	Ile	Val	Asn	Asn	Leu	Gly	Ser
	245				250					255				260
	Ala	Arg	Leu	Ile	Phe	Gln	Pro	Ile	Glu	Glu	Ser	Phe	Tyr	Ile
35				265						270				275
	Ala	Lys	Val	Leu	Glu	Arg	Gly	Lys	Asp	Ala	Thr	Leu	Gln	Lys
			280					285					290	
	Asp	Val	Ala	Val	Ala	Ala	Ala	Val	Leu	Glu	Ser	Leu	Leu	Lys
		295					300					305		
40	Leu	Leu	Ala	Gly	Leu	Thr	Ile	Thr	Val	Phe	Gly	Phe	Ala	Tyr
		310				315						320		
	Leu	Ala	Leu	Asp	Ile	Tyr	Gly	Gly	Thr	Met	Leu	Ser	Ser	Gly
	325				330					335				340
	Pro	Val	Leu	Leu	Arg	Ser	Tyr	Cys	Leu	Tyr	Val	Leu	Leu	Leu
45				345						350				355
	Asn	Gly	Val	Thr	Glu	Cys	Leu	Thr	Phe	Ala	Ala	Met	Ser	Lys
			360					365					370	
	Val	Asp	Arg	Tyr	Asn	Phe	Val	Met	Leu	Ala	Leu	Ser	Ser	Ser
		375					380					385		
50	Val	Leu	Ser	Tyr	Leu	Leu	Thr	Arg	Trp	Cys	Gly	Ser	Val	Gly
		390					395					400		
	Leu	Ala	Asn	Cys	Phe	Asn	Met	Gly	Ile	Arg	Ile	Thr	Gln	Ser
	405				410					415				420
	Phe	Ile	His	Arg	Tyr	Arg	Arg	Ser	Pro	His	Arg	Pro	Leu	Ala
55				425					430					435
	Leu	His	Leu	Ser	Pro	Val	Leu	Leu	Gly	Thr	Phe	Ala	Leu	Ser
			440					445					450	
	Val	Thr	Ala	Val	Ser	Glu	Val	Phe	Leu	Cys	Cys	Asp	Gln	Gly
		455					460					465		
60	Ala	Arg	Leu	Ala	His	Ile	Ala	Val	Gly	Ala	Phe	Cys	Leu	Gly
		470				475						480		
	Leu	Gly	Thr	Ala	Phe	Leu	Thr	Glu	Thr	Lys	Leu	Ile	His	Phe
	485					490				495				500

Thr Gln Leu Gly Val Pro Arg Arg Thr Asp Lys Met Thr
505 510

<210> 301
5 <211> 287
<212> PRT
<213> Homo sapiens

<220>
10 <221> SIGNAL
<222> -17...-1

<400> 301
Met Glu Leu Glu Arg Ile Val Ser Ala Ala Leu Leu Ala Phe Val Gln
15 -15 -10 -5
Thr His Leu Pro Glu Ala Asp Leu Ser Gly Leu Asp Glu Val Ile Phe
1 5 10 15
Ser Tyr Val Leu Gly Val Leu Glu Asp Leu Gly Pro Ser Gly Pro Ser
20 20 25 30
Glu Glu Asn Phe Asp Met Glu Ala Phe Thr Glu Met Met Glu Ala Tyr
35 40 45
Val Pro Gly Phe Ala His Ile Pro Arg Gly Thr Ile Gly Asp Met Met
50 55 60
Gln Lys Leu Ser Gly Gln Leu Ser Asp Ala Arg Asn Lys Glu Asn Leu
25 65 70 75
Gln Pro Gln Ser Ser Gly Val Gln Gly Gln Val Pro Ile Ser Pro Glu
80 85 90 95
Pro Leu Gln Arg Pro Glu Met Leu Lys Glu Thr Arg Ser Ser Ala
100 105 110
30 Ala Ala Ala Ala Asp Thr Gln Asp Glu Ala Thr Gly Ala Glu Glu Glu
115 120 125
Leu Leu Pro Gly Val Asp Val Leu Leu Glu Val Phe Pro Thr Cys Ser
130 135 140
Val Glu Gln Ala Gln Trp Val Leu Ala Lys Ala Arg Gly Asp Leu Glu
35 145 150 155
Glu Ala Val Gln Met Leu Val Glu Gly Lys Glu Gly Pro Ala Ala
160 165 170 175
Trp Glu Gly Pro Asn Gln Asp Leu Pro Arg Arg Leu Arg Gly Pro Gln
180 185 190
40 Lys Asp Glu Leu Lys Ser Phe Ile Leu Gln Lys Tyr Met Met Val Asp
195 200 205
Ser Ala Glu Asp Gln Lys Ile His Arg Pro Met Ala Pro Lys Glu Ala
210 215 220
Pro Lys Lys Leu Ile Arg Tyr Ile Asp Asn Gln Val Val Ser Thr Lys
45 225 230 235
Gly Glu Arg Phe Lys Asp Val Arg Asn Pro Glu Ala Glu Glu Met Lys
240 245 250 255
Ala Thr Tyr Ile Asn Leu Lys Pro Ala Arg Lys Tyr Arg Phe His
260 265 270

50
<210> 302
<211> 165
<212> PRT
<213> Homo sapiens

55
<220>
<221> SIGNAL
<222> -35...-1

60 <400> 302
Met Met Arg Cys Cys Arg Arg Arg Cys Cys Cys Arg Gln Pro Pro His
-35 -30 -25 -20
Ala Leu Arg Pro Leu Leu Leu Leu Pro Leu Val Leu Leu Pro Pro Leu

-15 -10 -5
 Ala Ala Ala Ala Gly Pro Asn Arg Cys Asp Thr Ile Tyr Gln Gly
 1 5 10
 Phe Ala Glu Cys Leu Ile Arg Leu Gly Asp Ser Met Gly Arg Gly Gly
 15 20 25
 5 Glu Leu Glu Thr Ile Cys Arg Ser Trp Asn Tyr Phe His Ala Cys Ala
 30 35 40 45
 Ser Gln Val Leu Ser Gly Cys Pro Glu Glu Ala Ala Val Trp Glu
 50 55 60
 10 Ser Leu Gln Gln Glu Ala Arg Gln Ala Pro Arg Pro Asn Asn Leu His
 65 70 75
 Thr Leu Cys Gly Ala Pro Val His Val Arg Glu Arg Gly Thr Gly Ser
 80 85 90
 Glu Thr Asn Gln Glu Thr Leu Arg Ala Thr Ala Pro Ala Leu Pro Met
 95 100 105
 15 Ala Pro Ala Pro Pro Leu Leu Ala Ala Ala Leu Ala Leu Ala Tyr Leu
 110 115 120 125
 Leu Arg Pro Leu Ala
 130
 20
 <210> 303
 <211> 148
 <212> PRT
 <213> Homo sapiens
 25
 <220>
 <221> SIGNAL
 <222> -25...-1
 30 <400> 303
 Met Ala Ser Val Val Leu Ala Leu Arg Thr Arg Thr Ala Val Thr Ser
 -25 -20 -15 -10
 Leu Leu Ser Pro Thr Pro Ala Thr Ala Leu Ala Val Arg Tyr Ala Ser
 -5 1 5
 35 Lys Lys Ser Gly Gly Ser Ser Lys Asn Leu Gly Gly Lys Ser Ser Gly
 10 15 20
 Arg Arg Gln Gly Ile Lys Lys Met Glu Gly His Tyr Val His Ala Gly
 25 30 35
 Asn Ile Ile Ala Thr Gln Arg His Phe Arg Trp His Pro Gly Ala His
 40 40 45 50 55
 Val Gly Val Gly Lys Asn Lys Cys Leu Tyr Ala Leu Glu Glu Gly Ile
 60 65 70
 Val Arg Tyr Thr Lys Glu Val Tyr Val Pro His Pro Arg Asn Thr Glu
 75 80 85
 45 Ala Val Asp Leu Ile Thr Arg Leu Pro Lys Gly Ala Val Leu Tyr Lys
 90 95 100
 Thr Phe Val His Val Val Pro Ala Lys Pro Glu Gly Thr Phe Lys Leu
 105 110 115
 Val Ala Met Leu
 50 120
 <210> 304
 <211> 291
 <212> PRT
 55 <213> Homo sapiens
 <220>
 <221> SIGNAL
 <222> -34...-1
 60
 <400> 304
 Met Glu Ser Glu Arg Ser Lys Arg Met Gly Asn Ala Cys Ile Pro Leu
 -30 -25 -20

Lys Arg Ile Ala Tyr Phe Leu Cys Leu Leu Ser Ala Leu Leu Leu Thr
 -15. -10 -5
 Glu Gly Lys Lys Pro Ala Lys Pro Lys Cys Pro Ala Val Cys Thr Cys
 1 5 10
 5 Thr Lys Asp Asn Ala Leu Cys Glu Asn Ala Arg Ser Ile Pro Arg Thr
 15 20 25 30
 Val Pro Pro Asp Val Ile Ser Leu Ser Phe Val Arg Ser Gly Phe Thr
 35 40 45
 Glu Ile Ser Glu Gly Ser Phe Leu Phe Thr Pro Ser Leu Gln Leu Leu
 50 55 60
 10 Leu Phe Thr Ser Asn Ser Phe Asp Val Ile Ser Asp Asp Ala Phe Ile
 65 70 75
 Gly Leu Pro His Leu Glu Tyr Leu Phe Ile Glu Asn Asn Asn Ile Lys
 80 85 90
 15 Ser Ile Ser Arg His Thr Phe Arg Gly Leu Lys Ser Leu Ile His Leu
 95 100 105 110
 Ser Leu Ala Asn Asn Asn Leu Gln Thr Leu Pro Lys Asp Ile Phe Lys
 115 120 125
 20 Gly Leu Asp Ser Leu Thr Asn Val Asp Leu Arg Gly Asn Ser Phe Asn
 130 135 140
 Cys Asp Cys Lys Leu Lys Trp Leu Val Glu Trp Leu Gly His Thr Asn
 145 150 155
 Ala Thr Val Glu Asp Ile Tyr Cys Glu Gly Pro Pro Glu Tyr Lys Lys
 160 165 170
 25 Arg Lys Ile Asn Ser Leu Ser Ser Lys Asp Phe Asp Cys Ile Ile Thr
 175 180 185 190
 Glu Phe Ala Lys Ser Gln Asp Leu Pro Tyr Gln Ser Leu Ser Ile Asp
 195 200 205
 30 Thr Phe Ser Tyr Leu Asn Asp Glu Tyr Val Val Ile Ala Gln Pro Phe
 210 215 220
 Thr Gly Lys Cys Ile Phe Leu Glu Trp Asp His Val Glu Lys Thr Phe
 225 230 235
 Arg Asn Tyr Asp Asn Ile Thr Val Leu Arg Glu Ile His Arg Phe Thr
 240 245 250
 35 Asn Met Ser
 255

 <210> 305
 <211> 81
 40 <212> PRT
 <213> Homo sapiens

 <220>
 <221> SIGNAL
 45 <222> -49...-1

 <400> 305
 Met Glu Gly Ala Gly Ala Gly Ser Gly Phe Arg Lys Glu Leu Val Ser
 -45 -40 -35
 50 Arg Leu Leu His Leu His Phe Lys Asp Asp Lys Thr Lys Val Ser Gly
 -30 -25 -20
 Asp Ala Leu Gln Leu Met Val Glu Leu Leu Lys Val Phe Val Val Glu
 -15 -10 -5
 Ala Ala Val Arg Gly Val Arg Gln Ala Gln Ala Glu Asp Ala Leu Arg
 1 5 10 15
 55 Val Asp Val Asp Gln Leu Glu Lys Val Leu Pro Gln Leu Leu Leu Asp
 20 25 30
 Phe

 60 <210> 306
 <211> 233
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -30...-1
 5
 <400> 306
 Met Ala Ala Thr Ser Gly Thr Asp Glu Pro Val Ser Gly Glu Leu Val
 -30 -25 -20 -15
 Ser Val Ala His Ala Leu Ser Leu Pro Ala Glu Ser Tyr Gly Asn Asp
 10 -10 -5 1
 Pro Asp Ile Glu Met Ala Trp Ala Met Arg Ala Met Gln His Ala Glu
 5 10 15
 Val Tyr Tyr Lys Leu Ile Ser Ser Val Asp Pro Gln Phe Leu Lys Leu
 20 25 30
 Thr Lys Val Asp Asp Gln Ile Tyr Ser Glu Phe Arg Lys Asn Phe Glu
 15 35 40 45 50
 Thr Leu Arg Ile Asp Val Leu Asp Pro Glu Glu Leu Lys Ser Glu Ser
 55 60 65
 Ala Lys Glu Lys Trp Arg Pro Phe Cys Leu Lys Phe Asn Gly Ile Val
 20 70 75 80
 Glu Asp Phe Asn Tyr Gly Thr Leu Arg Leu Asp Cys Ser Gln Gly
 85 90 95
 Tyr Thr Glu Glu Asn Thr Ile Phe Ala Pro Arg Ile Gln Phe Phe Ala
 100 105 110
 Ile Glu Ile Ala Arg Asn Arg Glu Gly Tyr Asn Lys Ala Val Tyr Ile
 25 115 120 125 130
 Ser Val Gln Asp Lys Glu Gly Glu Lys Gly Val Asn Asn Gly Gly Glu
 135 140 145
 Lys Arg Ala Asp Ser Gly Glu Glu Glu Asn Thr Lys Asn Gly Gly Glu
 30 150 155 160
 Lys Gly Ala Asp Ser Gly Glu Glu Lys Glu Glu Gly Ile Asn Arg Glu
 165 170 175
 Asp Lys Thr Asp Lys Gly Gly Glu Lys Gly Lys Glu Ala Asp Lys Glu
 180 185 190
 35 Ile Asn Lys Ser Gly Glu Lys Ala Met
 195 200

 <210> 307
 <211> 85
 40 <212> PRT
 <213> Homo sapiens

 <220>
 <221> SIGNAL
 45 <222> -20...-1

 <400> 307
 Met Arg Gln Lys Ala Val Ser Leu Phe Leu Cys Tyr Leu Leu Leu Phe
 -20 -15 -10 -5
 50 Thr Cys Ser Gly Val Glu Ala Gly Lys Lys Lys Cys Ser Glu Ser Ser
 1 5 10
 Asp Ser Gly Ser Gly Phe Trp Lys Ala Leu Thr Phe Met Ala Val Gly
 15 20 25
 Gly Gly Leu Ala Val Ala Gly Leu Pro Ala Leu Gly Phe Thr Gly Ala
 55 30 35 40
 Gly Ile Ala Ala Asn Ser Val Ala Ala Ser Leu Met Ser Trp Ser Ala
 45 50 55 60
 Ile Leu Asn Gly Gly
 65
 60
 <210> 308
 <211> 105
 <212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

5 <222> -43...-1

<400> 308

```

Met Gly Phe Thr Gly Ala Gly Ile Ala Ala Ser Ser Ile Ala Ala Lys
      -40      -35      -30
10 Met Met Ser Ala Ala Ala Ile Ala Asn Gly Gly Gly Val Ser Ala Gly
      -25      -20      -15
    Ser Leu Val Ala Thr Leu Gln Ser Val Gly Ala Ala Gly Leu Ser Thr
      -10      -5      1      5
    Ser Ser Asn Ile Leu Leu Ala Ser Val Gly Ser Val Leu Gly Ala Cys
      10      15      20
15 Leu Gly Asn Ser Pro Ser Ser Ser Leu Pro Ala Glu Pro Glu Ala Lys
      25      30      35
    Glu Asp Glu Ala Arg Glu Asn Val Pro Gln Gly Glu Pro Pro Lys Pro
      40      45      50
20 Pro Leu Lys Ser Glu Lys His Glu Glu
      55      60

```

<210> 309

<211> 291

25 <212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

30 <222> -34...-1

<400> 309

```

Met Glu Ser Glu Arg Ser Lys Arg Met Gly Asn Ala Cys Ile Pro Leu
      -30      -25      -20
35 Lys Arg Ile Ala Tyr Phe Leu Cys Leu Leu Ser Ala Leu Leu Thr
      -15      -10      -5
    Glu Gly Lys Lys Pro Ala Lys Pro Lys Cys Pro Ala Val Cys Thr Cys
      1      5      10
    Thr Lys Asp Asn Ala Leu Cys Glu Asn Ala Arg Ser Ile Pro Arg Thr
      15      20      25      30
40 Val Pro Pro Asp Val Ile Ser Leu Ser Phe Val Arg Ser Val Phe Thr
      35      40      45
    Glu Ile Ser Glu Gly Ser Phe Leu Phe Thr Pro Ser Leu Gln Leu Leu
      50      55      60
45 Leu Phe Thr Ser Asn Ser Phe Asp Val Ile Ser Asp Asp Ala Phe Ile
      65      70      75
    Gly Leu Pro His Leu Glu Tyr Leu Phe Ile Glu Asn Asn Asn Ile Lys
      80      85      90
    Ser Ile Ser Arg His Thr Phe Arg Gly Leu Lys Ser Leu Ile His Leu
      95      100      105      110
50 Ser Leu Ala Asn Asn Asn Leu Gln Thr Leu Pro Lys Asp Ile Phe Lys
      115      120      125
    Gly Leu Asp Ser Leu Thr Asn Val Asp Leu Arg Gly Asn Ser Phe Asn
      130      135      140
55 Cys Asp Cys Lys Leu Lys Trp Leu Val Glu Trp Leu Gly His Thr Asn
      145      150      155
    Ala Thr Val Glu Asp Ile Tyr Cys Glu Gly Pro Pro Glu Tyr Lys Lys
      160      165      170
    Arg Lys Ile Asn Ser Leu Ser Ser Lys Asp Phe Asp Cys Ile Ile Thr
      175      180      185      190
60 Glu Phe Ala Lys Ser Gln Asp Leu Pro Tyr Gln Ser Leu Ser Ile Asp
      195      200      205
    Thr Phe Ser Tyr Leu Asn Asp Glu Tyr Val Val Ile Ala Gln Pro Phe

```

210 215 220
 Thr Gly Lys Cys Ile Phe Leu Glu Trp Asp His Val Glu Lys Thr Phe
 225 230 235
 Arg Asn Tyr Asp Asn Ile Thr Val Leu Arg Glu Ile His Arg Phe Thr
 5 240 245 250
 Asn Met Ser
 255

 <210> 310
 10 <211> 426
 <212> PRT
 <213> Homo sapiens

 <220>
 15 <221> SIGNAL
 <222> -28...-1

 <400> 310
 20 Met Ser Pro Ala Phe Arg Ala Met Asp Val Glu Pro Arg Ala Lys Gly
 -25 -20 -15
 Val Leu Leu Glu Pro Phe Val His Gln Val Gly Gly His Ser Cys Val
 -10 -5 1
 Leu Arg Phe Asn Glu Thr Thr Leu Cys Lys Pro Leu Val Pro Arg Glu
 5 10 15 20
 25 His Gln Phe Tyr Glu Thr Leu Pro Ala Glu Met Arg Lys Phe Thr Pro
 25 30 35
 Gln Tyr Lys Gly Val Val Ser Val Arg Phe Glu Glu Asp Glu Asp Arg
 40 45 50
 Asn Leu Cys Leu Ile Ala Tyr Pro Leu Lys Gly Asp His Gly Ile Val
 30 55 60 65
 Asp Ile Val Asp Asn Ser Asp Cys Glu Pro Lys Ser Lys Leu Leu Arg
 70 75 80
 Trp Thr Thr Asn Lys Lys His His Val Leu Glu Thr Glu Lys Thr Pro
 85 90 95 100
 35 Lys Asp Trp Val Arg Gln His Arg Lys Glu Glu Lys Met Lys Ser His
 105 110 115
 Lys Leu Glu Glu Phe Glu Trp Leu Lys Lys Ser Glu Val Leu Tyr
 120 125 130
 Tyr Thr Val Glu Lys Lys Gly Asn Ile Ser Ser Gln Leu Lys His Tyr
 40 135 140 145
 Asn Pro Trp Ser Met Lys Cys His Gln Gln Gln Leu Gln Arg Met Lys
 150 155 160
 Glu Asn Ala Lys His Arg Asn Gln Tyr Lys Phe Ile Leu Leu Glu Asn
 165 170 175 180
 45 Leu Thr Ser Arg Tyr Glu Val Pro Cys Val Leu Asp Leu Lys Met Gly
 185 190 195
 Thr Arg Gln His Gly Asp Asp Ala Ser Glu Glu Lys Ala Ala Asn Gln
 200 205 210
 Ile Arg Lys Cys Gln Gln Ser Thr Ser Ala Val Ile Gly Val Arg Val
 50 215 220 225
 Cys Gly Met Gln Val Tyr Gln Ala Gly Ser Gly Gln Leu Met Phe Met
 230 235 240
 Asn Lys Tyr His Gly Arg Lys Leu Ser Met Gln Gly Phe Lys Glu Ala
 245 250 255 260
 55 Leu Phe Gln Phe Phe His Asn Gly Arg Tyr Leu Arg Arg Glu Leu Leu
 265 270 275
 Gly Pro Val Leu Lys Lys Leu Thr Glu Leu Lys Ala Val Leu Glu Arg
 280 285 290
 Gln Glu Ser Tyr Arg Phe Tyr Ser Ser Ser Leu Leu Val Ile Tyr Asp
 60 295 300 305
 Gly Lys Glu Arg Pro Glu Val Val Leu Asp Ser Asp Ala Glu Asp Leu
 310 315 320
 Glu Asp Leu Ser Glu Glu Ser Ala Asp Glu Ser Ala Gly Ala Tyr Ala

325 330 335 340
 Tyr Lys Pro Ile Gly Ala Ser Ser Val Asp Val Arg Met Ile Asp Phe
 345 350 355
 Ala His Thr Thr Cys Arg Leu Tyr Gly Glu Asp Thr Val Val His Glu
 5 360 365 370
 Gly Gln Asp Ala Gly Tyr Ile Phe Gly Leu Gln Ser Leu Ile Asp Ile
 375 380 385
 Val Thr Glu Ile Ser Glu Glu Ser Gly Glu
 390 395
 10
 <210> 311
 <211> 466
 <212> PRT
 <213> Homo sapiens
 15
 <220>
 <221> SIGNAL
 <222> -16...-1
 20 <400> 311
 Met Gly Leu Tyr Ala Ala Ala Ala Gly Val Leu Ala Gly Val Glu Ser
 -15 -10 -5
 Arg Gln Gly Ser Ile Lys Gly Leu Val Tyr Ser Ser Asn Phe Gln Asn
 1 5 10 15
 25 Val Lys Gln Leu Tyr Ala Leu Val Cys Glu Thr Gln Arg Tyr Ser Ala
 20 25 30
 Val Leu Asp Ala Val Ile Ala Ser Ala Gly Leu Leu Arg Ala Glu Lys
 35 40 45
 Lys Leu Arg Pro His Leu Ala Lys Val Leu Val Tyr Glu Leu Leu Leu
 30 50 55 60
 Gly Lys Gly Phe Arg Gly Gly Gly Gly Arg Trp Lys Ala Leu Leu Gly
 65 70 75 80
 Arg His Gln Ala Arg Leu Lys Ala Glu Leu Ala Arg Leu Lys Val His
 85 90 95
 35 Arg Gly Val Ser Arg Asn Glu Asp Leu Leu Glu Val Gly Ser Arg Pro
 100 105 110
 Gly Pro Ala Ser Gln Leu Pro Arg Phe Val Arg Val Asn Thr Leu Lys
 115 120 125
 Thr Cys Ser Asp Asp Val Val Asp Tyr Phe Lys Arg Gln Gly Phe Ser
 40 130 135 140
 Tyr Gln Gly Arg Ala Ser Ser Leu Asp Asp Leu Arg Ala Leu Lys Gly
 145 150 155 160
 Lys His Phe Leu Leu Asp Pro Leu Met Pro Glu Leu Leu Val Phe Pro
 165 170 175
 45 Ala Gln Thr Asp Leu His Glu His Pro Leu Tyr Arg Ala Gly His Leu
 180 185 190
 Ile Leu Gln Asp Arg Ala Ser Cys Leu Pro Ala Met Leu Leu Asp Pro
 195 200 205
 Pro Pro Gly Ser His Val Ile Asp Ala Cys Ala Ala Pro Gly Asn Lys
 50 210 215 220
 Thr Ser His Leu Ala Ala Leu Leu Lys Asn Gln Gly Lys Ile Phe Ala
 225 230 235 240
 Phe Asp Leu Asp Ala Lys Arg Leu Ala Ser Met Ala Thr Leu Leu Ala
 245 250 255
 55 Arg Ala Gly Val Ser Cys Cys Glu Leu Ala Glu Glu Asp Phe Leu Ala
 260 265 270
 Val Ser Pro Ser Asp Pro Arg Tyr His Glu Val His Tyr Ile Leu Leu
 275 280 285
 Asp Pro Ser Cys Ser Gly Ser Gly Met Pro Ser Arg Gln Leu Glu Glu
 60 290 295 300
 Pro Gly Ala Gly Thr Pro Ser Pro Val Arg Leu His Ala Leu Ala Gly
 305 310 315 320
 Phe Gln Gln Arg Ala Leu Cys His Ala Leu Thr Phe Pro Ser Leu Gln

```

          325          330          335
Arg Leu Val Tyr Ser Thr Cys Ser Leu Cys Gln Glu Glu Asn Glu Asp
          340          345          350
Val Val Arg Asp Ala Leu Gln Gln Asn Pro Gly Ala Phe Arg Leu Ala
5   355          360          365
Pro Ala Leu Pro Ala Trp Pro His Arg Gly Leu Ser Thr Phe Pro Gly
          370          375          380
Ala Glu His Cys Leu Arg Ala Ser Pro Glu Thr Thr Leu Ser Ser Gly
385          390          395          400
10 Phe Phe Val Ala Val Ile Glu Arg Val Glu Val Pro Ser Ser Ala Ser
          405          410          415
Gln Ala Lys Ala Ser Ala Pro Glu Arg Thr Pro Ser Pro Ala Pro Lys
          420          425          430
Arg Lys Lys Arg Gln Gln Arg Ala Ala Ala Gly Ala Cys Thr Pro Pro
15   435          440          445
Cys Thr
    450

<210> 312
20 <211> 382
    <212> PRT
    <213> Homo sapiens

<220>
25 <221> SIGNAL
    <222> -16...-1

<400> 312
30 Met Gly Leu Tyr Ala Ala Ala Ala Gly Val Leu Ala Gly Val Glu Ser
    -15          -10          -5
Arg Gln Gly Ser Ile Lys Gly Leu Val Tyr Ser Ser Asn Phe Gln Asn
1   5          10          15
Val Lys Gln Leu Tyr Ala Leu Val Cys Glu Thr Gln Arg Tyr Ser Ala
          20          25          30
35 Val Leu Asp Ala Val Ile Ala Ser Ala Gly Leu Leu Arg Ala Glu Lys
          35          40          45
Lys Leu Arg Pro His Leu Ala Lys Val Leu Val Tyr Glu Leu Leu Leu
          50          55          60
Gly Lys Gly Phe Arg Gly Gly Gly Gly Arg Trp Lys Ala Leu Leu Gly
40 65          70          75          80
Arg His Gln Ala Arg Leu Lys Ala Glu Leu Ala Arg Leu Lys Val His
          85          90          95
Arg Gly Val Ser Arg Asn Glu Asp Leu Leu Glu Val Gly Ser Arg Pro
          100          105          110
45 Gly Pro Ala Ser Gln Leu Pro Arg Phe Val Arg Val Asn Thr Leu Lys
          115          120          125
Thr Cys Ser Asp Asp Val Val Asp Tyr Phe Lys Arg Gln Gly Phe Ser
          130          135          140
Tyr Gln Gly Arg Ala Ser Ser Leu Asp Asp Leu Arg Ala Leu Lys Gly
50 145          150          155          160
Lys His Phe Leu Leu Asp Pro Leu Met Pro Glu Leu Leu Val Phe Pro
          165          170          175
Ala Gln Thr Asp Leu His Glu His Pro Leu Tyr Arg Ala Gly His Leu
          180          185          190
55 Ile Leu Gln Asp Arg Ala Ser Cys Leu Pro Ala Met Leu Leu Asp Pro
          195          200          205
Pro Pro Gly Ser His Val Ile Asp Ala Cys Ala Ala Pro Gly Asn Lys
          210          215          220
Thr Ser His Leu Ala Ala Leu Leu Lys Asn Gln Gly Lys Ile Phe Ala
60 225          230          235          240
Phe Asp Leu Asp Ala Lys Arg Leu Ala Ser Met Ala Thr Leu Leu Ala
          245          250          255
Arg Ala Gly Val Ser Cys Cys Glu Leu Ala Glu Glu Asp Phe Leu Ala

```

260 265 270
 Val Ser Pro Ser Asp Pro Arg Tyr His Glu Val His Tyr Ile Leu Leu
 275 280 285
 Asp Pro Ser Cys Ser Gly Ser Gly Met Pro Ser Arg Gln Leu Glu Glu
 5 290 295 300
 Pro Gly Ala Gly Thr Pro Ser Pro Val Arg Leu His Ala Leu Ala Ala
 305 310 315 320
 Ser Ser Ser Glu Pro Cys Ala Thr Arg Ser Leu Ser Leu Pro Cys Ser
 325 330 335
 10 Gly Ser Ser Thr Pro Arg Ala Pro Ser Ala Arg Arg Arg Met Lys Thr
 340 345 350
 Trp Cys Glu Met Arg Cys Ser Arg Thr Arg Ala Pro Ser Gly
 355 360 365

 15 <210> 313
 <211> 258
 <212> PRT
 <213> Homo sapiens

 20 <220>
 <221> SIGNAL
 <222> -36...-1

 <400> 313
 25 Met Glu Glu Leu Gln Glu Pro Leu Arg Gly Glu Leu Arg Leu Cys Phe
 -35 -30 -25
 Thr Gln Ala Ala Arg Thr Ser Leu Leu Leu Leu Arg Leu Asn Asp Ala
 -20 -15 -10 -5
 Ala Leu Arg Ala Leu Gln Glu Cys Gln Arg Gln Gln Val Arg Pro Val
 30 1 5 10
 Ile Ala Phe Gln Gly His Arg Gly Tyr Leu Arg Leu Pro Gly Pro Gly
 15 20 25
 Trp Ser Cys Leu Phe Ser Phe Ile Val Ser Gln Cys Cys Gln Glu Gly
 30 35 40
 35 Ala Gly Gly Ser Leu Asp Leu Val Cys Gln Arg Phe Leu Arg Ser Gly
 45 50 55 60
 Pro Asn Ser Leu His Cys Leu Gly Ser Leu Arg Glu Arg Leu Ile Ile
 65 70 75
 Trp Ala Ala Met Asp Ser Ile Pro Ala Pro Ser Ser Val Gln Gly His
 40 80 85 90
 Asn Leu Thr Glu Asp Ala Arg His Pro Glu Ser Trp Gln Asn Thr Gly
 95 100 105
 Gly Tyr Ser Glu Gly Asp Ala Val Ser Gln Pro Gln Met Ala Leu Glu
 110 115 120
 45 Glu Val Ser Val Ser Asp Pro Leu Ala Ser Asn Gln Gly Gln Ser Leu
 125 130 135 140
 Pro Gly Ser Ser Arg Glu His Met Ala Gln Trp Glu Val Arg Ser Gln
 145 150 155
 Thr His Val Pro Asn Arg Glu Pro Val Gln Ala Leu Pro Ser Ser Ala
 50 160 165 170
 Ser Arg Lys Arg Leu Asp Lys Lys Arg Ser Val Pro Val Ala Thr Val
 175 180 185
 Glu Leu Glu Glu Lys Arg Phe Arg Thr Leu Pro Leu Val Pro Pro Pro
 190 195 200
 55 Thr Arg Pro Asp Gln Ser Gly Phe Thr Arg Gly Arg Arg Leu Gly Ala
 205 210 215 220
 Arg Arg

 <210> 314
 60 <211> 280
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -33...-1

5 <400> 314
 Met Lys Ser Cys Gly Ser Met Leu Gly Leu Trp Gly Gln Arg Leu Pro
 -30 -25 -20
 Ala Ala Trp Val Leu Leu Leu Leu Pro Phe Leu Pro Leu Leu Leu
 -15 -10 -5
 10 Ala Ala Pro Ala Pro His Arg Ala Ser Tyr Lys Pro Val Ile Val Val
 1 5 10 15
 His Gly Leu Phe Asp Ser Ser Tyr Ser Phe Arg His Leu Leu Glu Tyr
 20 25 30
 15 Ile Asn Glu Thr His Pro Gly Thr Val Val Thr Val Leu Asp Leu Phe
 35 40 45
 Asp Gly Arg Glu Ser Leu Arg Pro Leu Trp Glu Gln Val Gln Gly Phe
 50 55 60
 Arg Glu Ala Val Val Pro Ile Met Ala Lys Ala Pro Gln Gly Val His
 65 70 75
 20 Leu Ile Cys Tyr Ser Gln Gly Gly Leu Val Cys Arg Ala Leu Leu Ser
 80 85 90 95
 Val Met Asp Asp His Asn Val Asp Ser Phe Ile Ser Leu Ser Ser Pro
 100 105 110
 Gln Met Gly Gln Tyr Gly Asp Thr Asp Tyr Leu Lys Trp Leu Phe Pro
 115 120 125
 25 Thr Ser Met Arg Ser Asn Leu Tyr Arg Ile Cys Tyr Ser Pro Leu Ile
 130 135 140
 Asn Gly Glu Arg Asp His Pro Asn Ala Thr Val Trp Arg Lys Asn Phe
 145 150 155
 30 Leu Arg Val Gly His Leu Val Leu Ile Gly Gly Pro Asp Asp Gly Val
 160 165 170 175
 Ile Thr Pro Trp Gln Ser Ser Phe Phe Gly Phe Tyr Asp Ala Asn Glu
 180 185 190
 Thr Val Leu Glu Met Glu Glu Gln Leu Val Tyr Leu Arg Asp Ser Phe
 195 200 205
 35 Gly Leu Lys Thr Leu Leu Ala Arg Gly Ala Ile Val Arg Cys Pro Met
 210 215 220
 Ala Gly Ile Ser His Thr Ala Trp His Ser Asn Arg Thr Leu Tyr Glu
 225 230 235
 40 Thr Cys Ile Glu Pro Trp Leu Ser
 240 245

<210> 315
 <211> 174
 45 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 50 <222> -33...-1

<400> 315
 Met Lys Ser Cys Gly Ser Met Leu Gly Leu Trp Gly Gln Arg Leu Pro
 -30 -25 -20
 55 Ala Ala Trp Val Leu Leu Leu Leu Pro Phe Leu Pro Leu Leu Leu
 -15 -10 -5
 Ala Ala Pro Ala Pro His Arg Ala Ser Tyr Lys Pro Val Ile Val Val
 1 5 10 15
 His Gly Leu Phe Asp Ser Ser Tyr Ser Phe Arg His Leu Leu Glu Tyr
 20 25 30
 60 Ile Asn Glu Thr His Pro Gly Thr Val Val Thr Val Leu Asp Leu Phe
 35 40 45
 Asp Gly Arg Glu Ser Leu Arg Pro Leu Trp Glu Gln Val Gln Gly Phe

50 55 60
 Arg Glu Ala Val Val Pro Ile Met Ala Lys Ala Pro Gln Gly Val His
 65 70 75
 Leu Ile Cys Tyr Ser Gln Gly Gly Leu Val Cys Arg Ala Leu Leu Ser
 5 80 85 90 95
 Val Met Asp Asp His Asn Val Asp Ser Phe Ile Ser Leu Ser Ser Pro
 100 105 110
 Gln Met Gly Gln Tyr Gly Asp Thr Asp Tyr Leu Lys Trp Leu Phe Pro
 115 120 125
 10 Thr Ser Met Arg Ser Asn Leu Tyr Arg Ile Cys Tyr Ser Pro
 130 135 140

<210> 316

<211> 160

15 <212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

20 <222> -17...-1

<400> 316

Met Ala Phe Thr Phe Ala Ala Phe Cys Tyr Met Leu Ser Leu Val Leu
 -15 -10 -5
 25 Cys Ala Ala Leu Ile Phe Phe Ala Ile Trp His Ile Ile Ala Phe Asp
 1 5 10 15
 Glu Leu Arg Thr Asp Phe Lys Ser Pro Ile Asp Gln Cys Asn Pro Val
 20 25 30
 His Ala Arg Glu Arg Leu Arg Asn Ile Glu Arg Ile Cys Phe Leu Leu
 30 35 40 45
 Arg Lys Leu Val Leu Pro Glu Tyr Ser Ile His Ser Leu Phe Cys Ile
 50 55 60
 Met Phe Leu Cys Ala Gln Glu Trp Leu Thr Leu Gly Leu Asn Val Pro
 65 70 75
 35 Leu Leu Phe Tyr His Phe Trp Arg Tyr Phe His Cys Pro Ala Asp Ser
 80 85 90 95
 Ser Glu Leu Ala Tyr Asp Pro Pro Val Val Met Asn Pro Asp Thr Leu
 100 105 110
 Ser Tyr Cys Gln Lys Glu Ala Trp Cys Lys Leu Ala Phe Tyr Leu Leu
 40 115 120 125
 Ser Phe Phe Tyr Tyr Leu Tyr Cys Met Ile Tyr Thr Leu Val Ser Ser
 130 135 140

<210> 317

45 <211> 426

<212> PRT

<213> Homo sapiens

<220>

50 <221> SIGNAL

<222> -28...-1

<400> 317

Met Ser Pro Ala Phe Arg Ala Met Asp Val Glu Pro Arg Ala Lys Gly
 -25 -20 -15
 55 Val Leu Leu Glu Pro Phe Val His Gln Val Gly Gly His Ser Cys Val
 -10 -5 1
 Leu Arg Phe Asn Glu Thr Thr Leu Cys Lys Pro Leu Val Pro Arg Glu
 5 10 15 20
 60 His Gln Phe Tyr Glu Thr Leu Pro Ser Glu Met Arg Lys Phe Thr Pro
 25 30 35
 Gln Tyr Lys Gly Val Val Ser Val Arg Phe Glu Glu Asp Glu Asp Arg
 40 45 50

Asn Leu Cys Leu Ile Ala Tyr Pro Leu Lys Gly Asp His Gly Ile Val
 55 60 65
 Asp Ile Val Asp Asn Ser Asp Cys Glu Pro Lys Ser Lys Leu Leu Arg
 70 75 80
 5 Trp Thr Thr Asn Lys Lys His His Val Leu Glu Thr Glu Lys Thr Pro
 85 90 95 100
 Lys Asp Trp Val Arg Gln His Arg Lys Glu Glu Lys Met Lys Ser His
 105 110 115
 10 Lys Leu Glu Glu Glu Phe Glu Trp Leu Lys Lys Ser Glu Val Leu Tyr
 120 125 130
 Tyr Thr Val Glu Lys Lys Gly Asn Ile Ser Ser Gln Leu Lys His Tyr
 135 140 145
 Asn Pro Trp Ser Met Lys Cys His Gln Gln Gln Leu Gln Arg Met Lys
 150 155 160
 15 Glu Asn Ala Lys His Arg Asn Gln Tyr Lys Phe Ile Leu Leu Glu Asn
 165 170 175 180
 Leu Thr Ser Arg Tyr Glu Val Pro Cys Val Leu Asp Leu Lys Met Gly
 185 190 195
 20 Thr Arg Gln His Gly Asp Asp Ala Ser Glu Glu Lys Ala Ala Asn Gln
 200 205 210
 Ile Arg Lys Cys Gln Gln Ser Thr Ser Ala Val Ile Gly Val Arg Val
 215 220 225
 Cys Gly Met Gln Val Tyr Gln Ala Gly Ser Gly Gln Leu Met Phe Met
 230 235 240
 25 Asn Lys Tyr His Gly Arg Lys Leu Ser Val Gln Gly Phe Lys Glu Ala
 245 250 255 260
 Leu Phe Gln Phe Phe His Asn Gly Arg Tyr Leu Arg Arg Glu Leu Leu
 265 270 275
 30 Gly Pro Val Leu Lys Lys Leu Thr Glu Leu Lys Ala Val Leu Glu Arg
 280 285 290
 Gln Glu Ser Tyr Arg Phe Tyr Ser Ser Ser Leu Leu Val Ile Tyr Asp
 295 300 305
 Gly Lys Glu Arg Pro Glu Val Val Leu Asp Ser Asp Ala Glu Asp Leu
 310 315 320
 35 Glu Asp Leu Ser Glu Glu Ser Ala Asp Glu Ser Ala Gly Ala Tyr Ala
 325 330 335 340
 Tyr Lys Pro Ile Gly Ala Ser Ser Val Asp Val Arg Met Ile Asp Phe
 345 350 355
 40 Ala His Thr Thr Cys Arg Leu Tyr Gly Glu Asp Thr Val Val His Glu
 360 365 370
 Gly Gln Asp Ala Gly Tyr Ile Phe Gly Leu Gln Ser Leu Ile Asp Ile
 375 380 385
 Val Thr Glu Ile Ser Glu Glu Ser Gly Glu
 390 395
 45
 <210> 318
 <211> 301
 <212> PRT
 <213> Homo sapiens
 50
 <220>
 <221> SIGNAL
 <222> -20..-1
 55 <400> 318
 Met Ala Arg His Gly Leu Pro Leu Leu Pro Leu Leu Ser Leu Leu Val
 -20 -15 -10 -5
 Gly Ala Trp Leu Lys Leu Gly Asn Gly Gln Ala Thr Ser Met Val Gln
 1 5 10
 60 Leu Gln Gly Gly Arg Phe Leu Met Gly Thr Asn Ser Pro Asp Ser Arg
 15 20 25
 Asp Gly Glu Gly Pro Val Arg Glu Ala Thr Val Lys Pro Phe Ala Ile
 30 35 40

Asp Ile Phe Pro Val Thr Asn Lys Asp Phe Arg Asp Phe Val Arg Glu
 45 50 55 60
 Lys Lys Tyr Arg Thr Glu Ala Glu Met Phe Gly Leu Ser Phe Val Phe
 65 70 75
 5 Glu Asp Phe Val Ser Asp Glu Leu Arg Asn Lys Ala Thr Gln Pro Met
 80 85 90
 Lys Ser Val Leu Trp Trp Leu Pro Val Glu Lys Ala Phe Trp Arg Gln
 95 100 105
 Pro Ala Gly Pro Gly Ser Gly Ile Arg Glu Arg Leu Glu His Pro Val
 110 115 120
 10 Leu His Val Ser Trp Asn Asp Ala Arg Ala Tyr Cys Ala Trp Arg Gly
 125 130 135 140
 Lys Arg Leu Pro Thr Glu Glu Glu Trp Glu Phe Ala Ala Arg Gly Gly
 145 150 155
 15 Leu Lys Gly Gln Val Tyr Pro Trp Gly Asn Trp Phe Gln Pro Asn Arg
 160 165 170
 Thr Asn Leu Trp Gln Gly Lys Phe Pro Lys Gly Asp Lys Ala Glu Asp
 175 180 185
 Gly Phe His Gly Val Ser Pro Val Asn Ala Phe Pro Ala Gln Asn Asn
 190 195 200
 20 Tyr Gly Leu Tyr Asp Leu Leu Gly Asn Val Trp Glu Trp Thr Ala Ser
 205 210 215 220
 Pro Tyr Gln Ala Ala Glu Gln Asp Met Arg Val Leu Arg Gly Ala Ser
 225 230 235
 25 Trp Ile Asp Thr Ala Asp Gly Ser Ala Asn His Arg Ala Arg Val Thr
 240 245 250
 Thr Arg Met Gly Asn Thr Pro Asp Ser Ala Ser Asp Asn Leu Gly Phe
 255 260 265
 Arg Cys Ala Ala Asp Ala Gly Arg Pro Pro Gly Glu Leu
 270 275 280

<210> 319
 <211> 119
 <212> PRT
 35 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -17...-1

40 <400> 319
 Met Gly Ser Gly Trp Leu Thr Ala Val Ala Ser Leu Leu Pro Ser Pro
 -15 -10 -5
 Gly Asn Ser Glu Leu Pro Val Gln Ala Leu Gly Arg Arg Gly Gly Arg
 45 1 5 10 15
 Asp Trp Ala Arg Asn Glu Ala Gly Arg Asp Leu Glu Lys Pro Pro Arg
 20 25 30
 Leu His Cys Ser Gly Arg Gly Arg Leu Glu Glu Pro Val Pro Pro Asn
 35 40 45
 50 His Leu Pro Val Gly Leu Ser Val Arg Gly Ser Gln Val Leu Ser Ser
 50 55 60
 Ala Gly Pro Arg Arg Cys Arg Leu Thr Gly Thr Arg Asn Pro Val Arg
 65 70 75
 Gly Pro Arg Arg Val Glu Gln Ile Ala Arg Gly Gly Pro Glu Ala Arg
 80 85 90 95
 55 Arg Gln Ala Gly Asp Ser Cys
 100

<210> 320
 60 <211> 95
 <212> PRT
 <213> Homo sapiens

Met Pro Pro Thr Arg Asp Pro Phe Gln Gln Pro Thr Leu Asp Asn Asp
 -40 -35 -30
 Asp Ser Tyr Leu Gly Glu Leu Arg Ala Ser Lys Val Leu Trp Phe Leu
 -25 -20 -15 -10
 5 Ala Gln Ile Pro Ser Arg Val Ala Gly Ser Leu Leu Ser Val Cys Val
 -5 1 5
 Met Ser Arg Asp Gly Asn Ile Lys Asp Ser Gly Glu Asp Thr Gln Ser
 10 10 15 20
 Gly Thr Arg Glu Val Cys Phe Leu Pro Ala Ser Leu Ser Pro Tyr Ser
 10 25 30 35
 Ser Arg Leu Thr Phe Gln Arg Arg Phe
 40 45

 <210> 323
 15 <211> 70
 <212> PRT
 <213> Homo sapiens

 <220>
 20 <221> SIGNAL
 <222> -38...-1

 <400> 323
 25 Met Ser Ser Pro Gln Leu Pro Ala Phe Leu Trp Asp Lys Gly Thr Leu
 -35 -30 -25
 Thr Thr Ala Ile Ser Asn Pro Ala Cys Leu Val Asn Val Leu Phe Phe
 -20 -15 -10
 Phe Thr Pro Leu Met Thr Leu Val Thr Leu Leu Ile Leu Val Trp Lys
 -5 1 5 10
 30 Val Thr Lys Asp Lys Ser Asn Lys Asn Arg Glu Thr His Pro Arg Lys
 15 20 25
 Glu Ala Thr Trp Leu Pro
 30

 35 <210> 324
 <211> 168
 <212> PRT
 <213> Homo sapiens

 40 <220>
 <221> SIGNAL
 <222> -25...-1

 <400> 324
 45 Met Arg Gly Pro Thr Ala Gly Pro Ser Val Leu Ser Ala Ala His Leu
 -25 -20 -15 -10
 Leu Val Val Ile Leu Pro Ala Asn Ala Ala Leu Lys Leu Leu Ser Trp
 -5 1 5
 Glu Arg Leu Ala Ala Pro Ala Ile Glu Val Glu Val Pro Ser Lys Glu
 10 15 20
 50 Val Leu Ala Ala Pro Thr Lys Ala Lys Leu Ile Pro Ser Glu Asp Met
 25 30 35
 Leu Ala Ala Pro Ala Met Asp Leu Leu Asp Ser Phe Ser Pro Gly Phe
 40 45 50 55
 55 Leu Ile Ala Ala Pro Ala Ser Ala Val Ile Thr Trp Pro Gly Pro Ala
 60 65 70
 Asp Leu Val Val Ala Met Leu Ile Ala Pro Val Ala Gly Leu Ile Ala
 75 80 85
 Ala Pro Ala Ile Ala Thr Ser Val Leu Gly Pro Val Ala Val Pro Ala
 90 95 100
 60 Thr Ala Met Pro Pro Ala Val Leu Ala Ala Pro Pro Ser Ala Ala Pro
 105 110 115
 Gly Val Leu Val Asp Gly Glu Ala Ala Leu Ala Val Pro Trp Glu Ala

120 125 130 135
 Cys Trp Ile Pro Ser Pro Pro Ala
 140

5 <210> 325
 <211> 166
 <212> PRT
 <213> Homo sapiens

10 <220>
 <221> SIGNAL
 <222> -15...-1

<400> 325
 15 Met Leu Pro Leu Leu Ile Ile Cys Leu Leu Pro Ala Ile Glu Gly Lys
 -15 -10 -5 1
 Asn Cys Leu Arg Cys Trp Pro Glu Leu Ser Ala Leu Ile Asp Tyr Asp
 5 10 15
 Leu Gln Ile Leu Trp Val Thr Pro Gly Pro Pro Thr Glu Leu Ser Gln
 20 20 25 30
 Asn Arg Asp His Leu Glu Glu Glu Thr Ala Lys Phe Phe Thr Gln Val
 35 40 45
 His Gln Ala Ile Lys Thr Leu Arg Asp Asp Lys Thr Val Leu Leu Glu
 50 55 60 65
 25 Glu Ile Tyr Thr His Lys Asn Leu Phe Thr Glu Arg Leu Asn Lys Ile
 70 75 80
 Ser Asp Gly Leu Lys Glu Lys Asp Ile Gln Ser Thr Leu Lys Val Thr
 85 90 95
 Ser Cys Ala Asp Cys Arg Thr His Phe Leu Ser Cys Asn Asp Pro Thr
 100 105 110
 30 Phe Cys Pro Ala Arg Asn Arg Arg Thr Ser Leu Trp Ala Val Ser Leu
 115 120 125
 Ser Ser Ala Leu Leu Leu Ala Ile Ala Gly Asp Val Ser Phe Thr Gly
 130 135 140 145
 35 Lys Gly Arg Arg Arg Gln
 150

<210> 326
 <211> 156
 40 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 45 <222> -15...-1

<400> 326
 Met Asn Ile Leu Met Leu Thr Phe Ile Ile Cys Gly Leu Leu Thr Arg
 -15 -10 -5 1
 50 Val Thr Lys Gly Ser Phe Glu Pro Gln Lys Cys Trp Lys Asn Asn Val
 5 10 15
 Gly His Cys Arg Arg Arg Cys Leu Asp Thr Glu Arg Tyr Ile Leu Leu
 20 25 30
 Cys Arg Asn Lys Leu Ser Cys Cys Ile Ser Ile Ile Ser His Glu Tyr
 35 40 45
 55 Thr Arg Arg Pro Ala Phe Pro Val Ile His Leu Glu Asp Ile Thr Leu
 50 55 60 65
 Asp Tyr Ser Asp Val Asp Ser Phe Thr Gly Ser Pro Val Ser Met Leu
 70 75 80
 60 Asn Asp Leu Ile Thr Phe Asp Thr Thr Lys Phe Gly Glu Thr Met Thr
 85 90 95
 Pro Glu Thr Asn Thr Pro Glu Thr Met Pro Pro Ser Glu Ala Thr
 100 105 110

Thr Pro Glu Thr Thr Met Pro Pro Ser Glu Thr Ala Thr Ser Glu Thr
 115 120 125
 Met Pro Pro Pro Ser Gln Thr Ala Leu Thr His Asn
 130 135 140
 5
 <210> 327
 <211> 105
 <212> PRT
 <213> Homo sapiens
 10
 <220>
 <221> SIGNAL
 <222> -32...-1
 15
 <400> 327
 Met Ala Lys Met Phe Asp Leu Arg Thr Lys Ile Met Ile Gly Ile Glu
 -30 -25 -20
 Ser Ser Leu Leu Val Ala Ala Met Val Leu Leu Ser Val Val Phe Cys
 -15 -10 -5
 20 Leu Tyr Phe Lys Val Ala Lys Ala Leu Lys Ala Ala Lys Asp Pro Asp
 1 5 10 15
 Ala Val Ala Val Lys Asn His Asn Pro Asp Lys Val Cys Trp Ala Thr
 20 25 30
 Asn Ser Gln Ala Lys Ala Thr Thr Met Glu Ser Cys Pro Ser Leu Gln
 25 35 40 45
 Cys Cys Glu Gly Cys Arg Met His Ala Ser Ser Asp Ser Leu Pro Pro
 50 55 60
 Cys Cys Cys Asp Ile Asn Glu Gly Leu
 65 70
 30
 <210> 328
 <211> 81
 <212> PRT
 <213> Homo sapiens
 35
 <220>
 <221> SIGNAL
 <222> -27...-1
 40
 <400> 328
 Met Ser Asp Glu Asp Glu Ser Ser Asp Tyr Leu Cys Leu Ser Ile Leu
 -25 -20 -15
 Gly Leu Phe Cys Cys Leu Pro Leu Ala Ile Pro Ala Val Ile Phe Ser
 -10 -5 1 5
 45 Cys Leu Thr Lys Asn Tyr Asn Lys Ser Ser Asp Tyr Glu Leu Ala Ala
 10 15 20
 Lys Thr Ser Lys Gln Ala Tyr Tyr Trp Ala Ile Ala Ser Ile Thr Val
 25 30 35
 Gly Ile Leu Gly Thr Ile Leu Tyr Thr Tyr Leu Ile Tyr Leu Leu Arg
 50 40 45 50
 Leu
 <210> 329
 <211> 95
 55 <212> PRT
 <213> Homo sapiens
 <220>
 <221> SIGNAL
 60 <222> -27...-1
 <400> 329
 Met Thr Asp Gln Asp Arg Ile Ile Asn Leu Val Val Gly Ser Leu Thr

```

      -25      -20      -15
Ser Leu Leu Ile Leu Val Thr Leu Ile Ser Ala Phe Val Phe Pro Gln
  -10      -5      1      5
Leu Pro Pro Lys Pro Leu Asn Ile Phe Phe Ala Val Cys Ile Ser Leu
5      10      15      20
Ser Ser Ile Thr Ala Cys Ile Ile Tyr Trp Tyr Arg Gln Gly Asp Leu
      25      30      35
Glu Pro Lys Phe Arg Lys Leu Ile Tyr Tyr Ile Ile Phe Ser Ile Ile
      40      45      50
10 Met Leu Cys Ile Cys Ala Asn Leu Tyr Phe His Asp Val Gly Arg
      55      60      65

```

<210> 330

<211> 84

15 <212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

20 <222> -20...-1

<400> 330

```

Met Ala Ala Ala Ala Val Pro Ser Leu Leu Leu Ser Leu Pro Pro His
-20      -15      -10      -5
25 Gln Gly Leu Thr Phe Ser Asn Lys Ile Gln Pro Phe Gly Ala Gln Gly
      1      5      10
Val Leu His Pro Glu Pro Gly Leu Arg Asp Trp Leu Leu Pro Thr Cys
      15      20      25
Ser Arg Gln Leu Arg Val Ala Leu Pro Glu Lys Gly Ser Glu Gly Ser
30      30      35      40
Leu Cys Gln Thr Gln Leu Pro Ala Thr Pro Cys Phe Leu Pro Ser Asn
45      50      55      60
Thr Val Arg Thr

```

35 <210> 331

<211> 124

<212> PRT

<213> Homo sapiens

40 <220>

<221> SIGNAL

<222> -32...-1

<400> 331

```

45 Met Val Val Val Glu Pro Gly Ala Ser Leu Phe Pro Asn Gly Val Pro
      -30      -25      -20
Trp Leu Tyr Ala Val Phe Ala Val Leu Phe Val Phe Phe Leu Phe Ala
      -15      -10      -5
Met Leu Ser Pro Phe Leu Leu Glu Ile Asp Gln His Ile Lys Lys Phe
50 1      5      10      15
Leu Ile Arg Cys Arg Tyr Ser Leu His Asn Thr Val His Lys Asp Lys
      20      25      30
Lys Asn Ser Glu Ile Lys Met Asp His Leu Glu Arg Pro Gly Cys Pro
      35      40      45
55 Leu Glu Ser Pro Arg Arg Gly Val Leu Gly Gly Lys Lys Asn Gly Met
      50      55      60
Gly Asn Asp Pro Leu Leu Phe Val Lys Val Thr Lys Glu Pro Arg Asp
65      70      75      80
Ser Glu Ala Glu Ile Tyr Thr Pro Gly Pro Ser Val
60      85      90

```

<210> 332

<211> 62

<212> PRT
 <213> Homo sapiens

<220>
 5 <221> SIGNAL
 <222> -46...-1

<400> 332
 10 Met Asp Gln Leu Val Phe Lys Glu Thr Ile Trp Asn Asp Ala Phe Trp
 -45 -40 -35
 Gln Asn Pro Trp Asp Gln Gly Gly Leu Ala Val Ile Ile Leu Phe Ile
 -30 -25 -20 -15
 Thr Ala Val Leu Leu Leu Ile Leu Phe Ala Ile Val Phe Gly Leu Leu
 -10 -5 1
 15 Thr Ser Thr Glu Asn Thr Gln Cys Glu Ala Gly Glu Glu Glu
 5 10 15

<210> 333
 <211> 150
 20 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 25 <222> -23...-1

<400> 333
 30 Met Ser Asn Gln Arg Leu Pro Leu Ile Phe Ser Leu Leu Phe Ile Cys
 -20 -15 -10
 Phe Phe Gly Glu Ser Phe Cys Ile Cys Asp Gly Thr Val Trp Thr Lys
 -5 1 5
 Val Gly Trp Glu Ile Leu Pro Glu Glu Val His Tyr Trp Lys Gly Cys
 10 15 20 25
 Leu Tyr Leu Ile Tyr Asn Leu Leu Gln Ala Val Phe Phe Val Leu Phe
 35 30 35 40
 Val Leu Ser Val His Tyr Leu Trp Lys Trp Lys Lys His Gln Lys
 45 50 55
 Lys Leu Lys Lys Gln Ala Ser Leu Glu Lys Pro Gly Asn Asp Leu Glu
 60 65 70
 40 Ser Pro Leu Ile Asn Asn Ile Asp Gln Thr Leu His Arg Val Ala Thr
 75 80 85
 Thr Ala Ser Val Ile Tyr Lys Ile Trp Glu His Arg Ser His His Pro
 90 95 100 105
 Ser Ser Lys Lys Ile Lys His Cys Lys Leu Lys Lys Lys Ser Lys Glu
 45 110 115 120
 Glu Gly Ala Arg Arg Tyr
 125

<210> 334
 50 <211> 198
 <212> PRT
 <213> Homo sapiens

<220>
 55 <221> SIGNAL
 <222> -13...-1

<400> 334
 60 Met Leu Leu Gly Arg Leu Thr Ser Gln Leu Leu Arg Ala Val Pro Trp
 -10 -5 1
 Ala Gly Gly Arg Pro Pro Trp Pro Val Ser Gly Val Leu Gly Ser Arg
 5 10 15
 Val Cys Gly Pro Leu Tyr Ser Thr Ser Pro Ala Gly Pro Gly Arg Ala

[illegible]

```

    <210> 335
    <211> 88
25  <212> PRT
    <213> Homo sapiens

```

```

    <220>
    <221> SIGNAL
30  <222> -24..-1

```

[illegible]

```
<210> 336
<211> 150
<212> PRT
<213> Homo sapiens
```

```
50      <220>
      <221>  SIGNAL
      <222> -45...-1
```

```

55  <400> 336
    Met Val Leu Met Trp Thr Ser Gly Asp Ala Phe Lys Thr Ala Tyr Phe
      -45              -40              -35              -30
    Leu Leu Lys Gly Ala Pro Leu Gln Phe Ser Val Cys Gly Leu Leu Gln
              -25              -20              -15
60  Val Leu Val Asp Leu Ala Ile Leu Gly Gln Ala Tyr Ala Phe Ala Pro
      -10              -5              1
    Pro Pro Glu Ala Gly Ala Pro Arg Arg Ala Pro His Trp His Gln Gly
      5              10              15

```


Pro Leu Thr Val Gly Arg Thr Arg Met Trp Asp Arg Gln Pro Arg Ala
 20 25 30 35
 Leu Val Gly Pro Asp Leu Pro Ala Gly Arg Val Gly Ala Val Ala Pro
 40 45 50
 5 Ala Gly Val Ala Glu Met Gly His Gly His Trp Gly Leu His Gln Pro
 55 60 65
 Leu Trp Gly Val Ser Gly Trp Ala Val Gly Val Gly Leu Gly Arg Cys
 70 75 80
 Leu Cys Ser Ala Gly Thr Ala Arg Val Asp Leu Ala Pro Arg Val Leu
 10 85 90 95
 Asp Val Phe Arg Met Thr
 100 105

<210> 337
 15 <211> 142
 <212> PRT
 <213> Homo sapiens

<220>
 20 <221> SIGNAL
 <222> -19...-1

<400> 337
 25 Met Ala Thr Ala Ser Pro Ser Val Phe Leu Leu Met Val Asn Gly Gln
 -15 -10 -5
 Val Glu Ser Ala Gln Phe Pro Glu Tyr Asp Asp Phe Tyr Cys Lys Tyr
 1 5 10
 Cys Phe Val Tyr Gly Gln Asp Trp Ala Pro Thr Ala Gly Leu Glu Glu
 15 20 25
 30 Gly Ile Ser Gln Ile Thr Ser Lys Ser Gln Asp Val Arg Gln Ala Leu
 30 35 40 45
 Val Trp Asn Phe Pro Ile Asp Val Thr Phe Lys Ser Thr Asn Pro Tyr
 50 55 60
 Gly Trp Pro Gln Ile Val Leu Ser Val Tyr Gly Pro Asp Val Phe Gly
 35 65 70 75
 Asn Asp Val Val Arg Gly Tyr Gly Ala Val His Val Pro Phe Ser Pro
 80 85 90
 Gly Arg His Lys Arg Thr Ile Pro Met Phe Val Pro Glu Ser Thr Ser
 95 100 105
 40 Lys Leu Gln Lys Phe Thr Arg Ser Ala Ser Cys Ser Thr His
 110 115 120

<210> 338
 <211> 112
 45 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 50 <222> -27...-1

<220>
 <221> UNSURE
 <222> 21
 55 <223> Xaa = Ala, Pro

<400> 338
 Thr Ser Glu Glu Arg Thr Ala Met Lys Arg Glu Gly Gly Ala Ala His
 -25 -20 -15
 60 Leu Cys Ser Asp Ser Leu Pro Glu Ser Gln Gln Gln Asp Gly Asn His
 -10 -5 1 5
 Ala Pro Asn Phe Ser Ser His Gly Ser Cys Arg Arg Arg Gln Arg Xaa
 10 15 20

Asp Met Thr Arg Arg Cys Met Pro Ala Arg Pro Gly Phe Pro Ser Ser
 25 30 35
 Pro Ala Pro Gly Ser Ser Pro Pro Arg Cys His Leu Arg Pro Gly Ser
 40 45 50
 5 Thr Ala His Ala Ala Ala Gly Lys Arg Thr Glu Ser Pro Gly Asp Arg
 55 60 65
 Tyr Arg Ala Glu Gly Leu Arg Arg Gly Arg Val Ala Gly Ala Arg Val
 70 75 80 85

10 <210> 339
 <211> 90
 <212> PRT
 <213> Homo sapiens

15 <220>
 <221> SIGNAL
 <222> -32...-1

<400> 339
 20 Met Pro Cys Leu Asp Gln Gln Leu Thr Val His Ala Leu Pro Cys Pro
 -30 -25 -20
 Ala Gln Pro Ser Ser Leu Ala Phe Cys Gln Val Gly Phe Leu Thr Ala
 -15 -10 -5
 Gln Pro Ser Pro Pro Arg Arg Arg Asn Gly Lys Asp Arg Tyr Thr Leu
 25 1 5 10 15
 Val Leu Gln His Gln Glu Cys Gln Asp Asp Leu Ala Thr Ser Ser Leu
 20 25 30
 Val Tyr Leu Ser Leu Pro Cys Phe Lys Asp Leu Gly Arg Ser Lys His
 35 40 45
 30 Gln Ser Ile Thr Val Ala Asp Thr Asn Lys
 50 55

<210> 340
 <211> 80
 35 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 40 <222> -35...-1

<400> 340
 Met Pro Phe Gln Phe Gly Thr Gln Pro Arg Arg Phe Pro Val Glu Gly
 -35 -30 -25 -20
 45 Gly Asp Ser Ser Ile Glu Leu Glu Pro Gly Leu Ser Ser Ser Ala Ala
 -15 -10 -5
 Cys Asn Gly Lys Glu Met Ser Pro Thr Arg Gln Leu Arg Arg Cys Pro
 1 5 10
 Gly Ser His Cys Leu Thr Ile Thr Asp Val Pro Val Thr Val Tyr Ala
 50 15 20 25
 Thr Thr Arg Lys Pro Pro Ala Gln Ser Ser Lys Glu Met His Pro Lys
 30 35 40 45

<210> 341
 55 <211> 131
 <212> PRT
 <213> Homo sapiens

<220>
 60 <221> SIGNAL
 <222> -15...-1

<400> 341

Met Ser Leu Leu Met Phe Thr Gln Leu Leu Leu Cys Gly Phe Leu Tyr
 -15 -10 -5 1
 Val Arg Val Asp Gly Ser Arg Leu Arg Gln Glu Asp Phe Pro Pro Arg
 5 5 10 15
 Ile Val Glu His Pro Ser Asp Val Ile Val Ser Lys Gly Glu Pro Thr
 20 25 30
 Thr Leu Asn Cys Lys Ala Glu Gly Arg Pro Thr Pro Thr Ile Glu Trp
 35 40 45
 Tyr Lys Asp Gly Glu Arg Val Glu Thr Asp Lys Asp Asp Pro Arg Ser
 10 50 55 60 65
 His Arg Met Leu Leu Pro Ser Gly Ser Leu Phe Phe Leu Arg Ile Val
 70 75 80
 His Gly Arg Arg Ser Lys Pro Asp Glu Gly Ser Tyr Val Cys Val Ala
 85 90 95
 15 Arg Asn Tyr Leu Gly Glu Ala Val Ser Arg Asn Ala Ser Leu Glu Val
 100 105 110
 Ala Cys Lys
 115

20 <210> 342
 <211> 99
 <212> PRT
 <213> Homo sapiens

25 <220>
 <221> SIGNAL
 <222> -39...-1

<400> 342
 30 Met Asp Leu Ile Gly Phe Gly Tyr Ala Ala Leu Val Thr Phe Gly Ser
 -35 -30 -25
 Ile Phe Gly Tyr Lys Arg Arg Gly Gly Val Pro Ser Leu Ile Ala Gly
 -20 -15 -10
 Leu Phe Val Gly Cys Leu Ala Gly Tyr Gly Ala Tyr Arg Val Ser Asn
 35 -5 1 5
 Asp Lys Arg Asp Val Lys Val Ser Leu Phe Thr Ala Phe Phe Leu Ala
 10 15 20 25
 Thr Ile Met Gly Val Arg Phe Lys Arg Ser Lys Lys Ile Met Pro Ala
 30 35 40
 40 Gly Leu Val Ala Gly Leu Ser Leu Met Met Ile Leu Arg Leu Val Leu
 45 50 55
 Leu Leu Leu
 60

45 <210> 343
 <211> 98
 <212> PRT
 <213> Homo sapiens

50 <220>
 <221> SIGNAL
 <222> -43...-1

<400> 343
 55 Met Cys Glu Thr Leu Leu Thr Ser Lys Trp Ala Ser Val Ser Pro Ile
 -40 -35 -30
 Pro Ala Leu Leu Gln Glu Gly Glu Asn Arg Asp Ser Arg Arg Leu Gly
 -25 -20 -15
 Asp Ala Leu Leu Phe Leu Arg Pro Ala Gly Ser Cys Ala Leu Gln Val
 60 -10 -5 1 5
 Ser Trp Pro Ala Ala Leu Ala Gly Pro Arg Ser His Thr Gly Gln Leu
 10 15 20
 Thr Gln His Phe Cys His Leu Lys Asn Asp Thr Cys Ile Pro Pro Ser

25 30 35
 Leu Gly Pro Pro Arg Asn Ser Gly Ser Leu Glu Ser Leu Arg Ser Lys
 40 45 50
 Arg Tyr
 5 55
 <210> 344
 <211> 217
 <212> PRT
 10 <213> Homo sapiens
 <220>
 <221> SIGNAL
 <222> -19...-1
 15
 <220>
 <221> UNSURE
 <222> 185
 <223> Xaa = Phe,Val
 20
 <400> 344
 Met Val Gly Ile Leu Pro Leu Cys Cys Ser Gly Cys Val Pro Ser Leu
 -15 -10 -5
 25 Cys Cys Ser Ser Tyr Val Pro Ser Val Ala Pro Thr Ala Ala His Ser
 1 5 10
 Val Arg Val Val Pro His Ser Ala Gly His Cys Gly Gln Arg Val Leu Ala
 15 20 25
 Cys Ser Leu Pro Gln Val Phe Leu Lys Pro Trp Ile Phe Val Glu His
 30 35 40 45
 Phe Ser Ser Trp Leu Ser Leu Glu Leu Phe Ser Phe Leu Arg Tyr Leu
 50 55 60
 Gly Thr Leu Leu Cys Ala Cys Gly His Arg Leu Arg Glu Gly Arg Leu
 65 70 75
 35 Leu Pro Cys Leu Leu Gly Val Gly Ser Trp Leu Leu Phe Asn Asn Trp
 80 85 90
 Thr Gly Gly Ser Trp Phe Ser Leu His Leu Gln Gln Val Ser Leu Ser
 95 100 105
 Gln Gly Ser His Val Ala Ala Phe Leu Pro Glu Ala Ile Gly Pro Gly
 110 115 120 125
 40 Val Pro Val Pro Val Ser Gly Glu Ser Thr Ser Ala Gln Gln Ser His
 130 135 140
 Ala Gly Trp Gln Leu Ser Ala Glu Ala Asp Ala Cys Pro Ser Val Leu
 145 150 155
 45 Tyr Ser Glu Val Leu Glu Trp Asn Lys Asn Ile Asn Thr Tyr Thr Ser
 160 165 170
 Phe His Asp Phe Cys Leu Ile Leu Gly Ile Phe Xaa Val Leu Phe Cys
 175 180 185
 Phe Gly Gly Asp Arg Leu Thr Leu His
 190 195
 50
 <210> 345
 <211> 183
 <212> PRT
 <213> Homo sapiens
 55
 <220>
 <221> SIGNAL
 <222> -20...-1
 60
 <400> 345
 Met Lys Leu Leu Ser Leu Val Ala Val Val Gly Cys Leu Leu Val Pro
 -20 -15 -10 -5
 Pro Ala Glu Ala Asn Lys Ser Ser Glu Asp Ile Arg Cys Lys Cys Ile

1 5 10
 Cys Pro Pro Tyr Arg Asn Ile Ser Gly His Ile Tyr Asn Gln Asn Val
 15 20 25
 Ser Gln Lys Asp Cys Asn Cys Leu His Val Val Glu Pro Met Pro Val
 30 35 40
 5 Pro Gly His Asp Val Glu Ala Tyr Cys Leu Leu Cys Glu Cys Arg Tyr
 45 50 55 60
 Glu Glu Arg Ser Thr Thr Thr Ile Lys Val Ile Ile Val Ile Tyr Leu
 65 70 75
 10 Ser Val Val Gly Ala Leu Leu Leu Tyr Met Ala Phe Leu Met Leu Val
 80 85 90
 Asp Pro Leu Ile Arg Lys Pro Asp Ala Tyr Thr Glu Gln Leu His Asn
 95 100 105
 Glu Glu Glu Asn Glu Asp Ala Arg Ser Met Ala Ala Ala Ala Ala Ser
 110 115 120
 15 Leu Gly Gly Pro Arg Ala Asn Thr Val Leu Glu Arg Val Glu Gly Ala
 125 130 135 140
 Gln Gln Arg Trp Lys Leu Gln Val Gln Glu Gln Arg Lys Thr Val Phe
 145 150 155
 20 Asp Arg His Lys Met Leu Ser
 160

<210> 346

<211> 247

25 <212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

30 <222> -13...-1

<400> 346

Met Leu Val Leu Arg Ser Ala Leu Thr Arg Ala Leu Ala Ser Arg Thr
 -10 -5 1
 35 Leu Ala Pro Gln Met Cys Ser Ser Phe Ala Thr Gly Pro Arg Gln Tyr
 5 10 15
 Asp Gly Ile Phe Tyr Glu Phe Arg Ser Tyr Tyr Leu Lys Pro Ser Lys
 20 25 30 35
 Met Asn Glu Phe Leu Glu Asn Phe Glu Lys Asn Ala His Leu Arg Thr
 40 45 50
 Ala His Ser Glu Leu Val Gly Tyr Trp Ser Val Glu Phe Gly Gly Arg
 55 60 65
 Met Asn Thr Val Phe His Ile Trp Lys Tyr Asp Asn Phe Ala His Arg
 70 75 80
 45 Thr Glu Val Gln Lys Ala Leu Ala Lys Asp Lys Glu Trp Gln Glu Gln
 85 90 95
 Phe Leu Ile Pro Asn Leu Ala Leu Ile Asp Lys Gln Glu Ser Glu Ile
 100 105 110 115
 Thr Tyr Leu Val Pro Trp Cys Lys Leu Glu Lys Pro Pro Lys Glu Gly
 120 125 130
 50 Val Tyr Glu Leu Ala Thr Phe Gln Met Lys Pro Gly Gly Pro Ala Leu
 135 140 145
 Trp Gly Asp Ala Phe Lys Arg Ala Val His Ala His Val Asn Leu Gly
 150 155 160
 55 Tyr Thr Lys Leu Val Gly Val Phe His Thr Glu Tyr Gly Ala Leu Asn
 165 170 175
 Arg Val His Val Leu Trp Trp Asn Glu Ser Ala Asp Ser Arg Ala Ala
 180 185 190 195
 Gly Arg His Lys Ser His Glu Asp Pro Arg Val Val Ala Ala Val Arg
 200 205 210
 60 Glu Ser Val Asn Tyr Leu Val Ser Gln Gln Asn Met Leu Leu Ile Pro
 215 220 225
 Thr Ser Phe Ser Pro Leu Lys

230

<210> 347
 <211> 104
 5 <212> PRT
 <213> Homo sapiens

 <220>
 <221> SIGNAL
 10 <222> -47...-1

 <400> 347
 Met Phe Ser Pro Arg Gln Ala Leu Thr Pro Asp Pro Leu His Ser Pro
 -45 -40 -35
 15 Ala Tyr Ser Pro Val Leu Gly Gly Trp Ser Arg Phe Arg Ser Val Asp
 -30 -25 -20
 Phe Arg Phe Leu Tyr Leu Thr Leu Asn Gln Ser Cys Ile Phe Ala Asn
 -15 -10 -5 1
 20 Tyr Lys Glu Ala His Ala Asn Arg Tyr Cys Thr Glu Gly Arg Tyr Thr
 5 10 15
 Arg Glu Ile Gln Arg Leu Thr Ser Pro Ala Ala Trp Pro Thr Arg Asp
 20 25 30
 Lys Asn Arg Met Ile Ser Asn Gly Met Ala Leu Asn Ser Pro Ala Glu
 35 40 45
 25 Gly Leu Ala Phe Gln Cys Arg Phe
 50 55

<210> 348
 <211> 125
 30 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 35 <222> -21...-1

<400> 348
 Met Ala Lys Tyr Leu Ala Gln Ile Ile Val Met Gly Val Gln Val Val
 -20 -15 -10
 40 Gly Arg Ala Phe Ala Arg Ala Leu Arg Gln Glu Phe Ala Ala Ser Arg
 -5 1 5 10
 Ala Ala Ala Asp Ala Arg Gly Arg Ala Gly His Arg Ser Ala Ala Ala
 15 20 25
 45 Ser Asn Leu Ser Gly Leu Ser Leu Gln Glu Ala Gln Gln Ile Leu Asn
 30 35 40
 Val Ser Lys Leu Ser Pro Glu Glu Val Gln Lys Asn Tyr Glu His Leu
 45 50 55
 Phe Lys Val Asn Asp Lys Ser Val Gly Gly Ser Phe Tyr Leu Gln Ser
 60 65 70 75
 50 Lys Val Val Arg Ala Lys Glu Arg Leu Asp Glu Glu Leu Lys Ile Gln
 80 85 90
 Ala Gln Glu Asp Arg Glu Lys Gly Gln Met Pro His Thr
 95 100

55 <210> 349
 <211> 302
 <212> PRT
 <213> Homo sapiens

60 <220>
 <221> SIGNAL
 <222> -18...-1

<400> 349
 Met Ala Pro Asn Ser Ile Thr Leu Leu Gly Leu Ala Val Asn Val Val
 -15 -10 -5
 5 Thr Thr Leu Val Leu Ile Ser Tyr Cys Pro Thr Ala Thr Glu Glu Ala
 1 5 10
 Pro Tyr Trp Thr Tyr Leu Leu Cys Ala Leu Gly Leu Phe Ile Tyr Gln
 15 20 25 30
 Ser Leu Asp Ala Ile Asp Gly Lys Gln Ala Arg Arg Thr Asn Ser Cys
 35 40 45
 10 Ser Pro Leu Gly Glu Leu Phe Asp His Gly Cys Asp Ser Leu Ser Thr
 50 55 60
 Val Phe Met Ala Val Gly Ala Ser Ile Ala Ala Arg Leu Gly Thr Tyr
 65 70 75
 15 Pro Asp Trp Phe Phe Phe Cys Ser Phe Ile Gly Met Phe Val Phe Tyr
 80 85 90
 Cys Ala His Trp Gln Thr Tyr Val Ser Gly Met Leu Arg Phe Gly Lys
 95 100 105 110
 Val Asp Val Thr Glu Ile Gln Ile Ala Leu Val Ile Val Phe Val Leu
 115 120 125
 20 Ser Ala Phe Gly Gly Ala Thr Met Trp Asp Tyr Thr Gly Thr Ser Val
 130 135 140
 Leu Ser Pro Gly Leu His Ile Gly Leu Ile Ile Ile Leu Ala Ile Met
 145 150 155
 25 Ile Tyr Lys Lys Ser Ala Thr Asp Val Phe Glu Lys His Pro Cys Leu
 160 165 170
 Tyr Ile Leu Met Phe Gly Cys Val Phe Ala Lys Val Ser Gln Lys Leu
 175 180 185 190
 Val Val Ala His Met Thr Lys Ser Glu Leu Tyr Leu Gln Asp Thr Val
 195 200 205
 30 Phe Leu Gly Pro Gly Leu Leu Phe Leu Asp Gln Tyr Phe Asn Asn Phe
 210 215 220
 Ile Asp Glu Tyr Val Val Leu Trp Met Ala Met Val Ile Ser Ser Phe
 225 230 235
 35 Asp Met Val Ile Tyr Phe Ser Ala Leu Cys Leu Gln Ile Ser Arg His
 240 245 250
 Leu His Leu Asn Ile Phe Lys Thr Ala Cys His Gln Ala Pro Glu Gln
 255 260 265 270
 Val Gln Val Leu Ser Ser Lys Ser His Gln Asn Asn Met Asp
 275 280

40
 <210> 350
 <211> 107
 <212> PRT
 <213> Homo sapiens

45
 <220>
 <221> SIGNAL
 <222> -14..-1

50 <400> 350
 Met Ile Leu Val Thr Val Pro Gly Val Cys Pro Ala Gln Cys Cys Trp
 -10 -5 1
 Ala Glu Gln Arg Gly Arg Gly Ser Gly Met Tyr Phe Ile Asp Lys Trp
 5 10 15
 55 Ala Arg Pro Ser Trp Val Pro His Trp Leu Asn Asp Leu Phe Ile Val
 20 25 30
 Lys Ser Gly Tyr Leu Val Cys Ile Arg Thr Thr Val Ile Arg Gln Gly
 35 40 45 50
 Ile Val Arg Ile Gly Arg Asn Lys Ile Ser Glu Ser Gly Arg Ser Ala
 55 60 65
 60 Leu Tyr Thr Ile Ala Lys Asn Lys Met Val Ile Phe Lys Val Pro Asp
 70 75 80
 Cys Met His Leu Asn Ala Asp Tyr Phe Gly Val

85

90

```

    <210> 351
    <211> 229
5  <212> PRT
    <213> Homo sapiens

```

```

    <220>
    <221> SIGNAL
10  <222> -34...-1

```

[illegible]

```

45  <210> 352
    <211> 206
    <212> PRT
    <213> Homo sapiens

```

```

    <220>
50  <221>  SIGNAL
    <222>  -34..-1

```

	<400>	352																
55	Met	Ser	Phe	Leu	Gln	Asp	Pro	Ser	Phe	Phe	Thr	Met	Gly	Met	Trp	Ser		
					-30					-25					-20			
	Ile	Gly	Ala	Gly	Ala	Leu	Gly	Ala	Ala	Ala	Leu	Ala	Leu	Leu	Leu	Ala		
				-15					-10					-5				
	Asn	Thr	Asp	Val	Phe	Leu	Ser	Lys	Pro	Gln	Lys	Ala	Ala	Leu	Glu	Tyr		
			1				5					10						
60	Leu	Glu	Asp	Ile	Asp	Leu	Lys	Thr	Leu	Glu	Lys	Glu	Pro	Arg	Thr	Phe		
	15					20					25					30		
	Lys	Ala	Lys	Glu	Leu	Trp	Glu	Lys	Asn	Gly	Ala	Val	Ile	Met	Ala	Val		
				35						40					45			

Arg Arg Pro Gly Cys Phe Leu Cys Arg Glu Glu Ala Ala Asp Leu Ser
 50 55 60
 Ser Leu Lys Ser Met Leu Asp Gln Leu Gly Val Pro Leu Tyr Ala Val
 65 70 75
 5 Val Lys Glu His Ile Arg Thr Glu Val Lys Asp Phe Gln Pro Tyr Phe
 80 85 90
 Lys Gly Glu Ile Phe Leu Asp Glu Lys Lys Lys Phe Tyr Gly Pro Gln
 95 100 105 110
 10 Arg Arg Lys Met Met Phe Met Gly Phe Ile Arg Leu Gly Val Trp Tyr
 115 120 125
 Asn Phe Phe Arg Ala Trp Asn Gly Gly Phe Ser Gly Asn Leu Glu Gly
 130 135 140
 Glu Gly Phe Ile Leu Gly Gly Val Phe Val Val Gly Ser Gly Ser Arg
 145 150 155
 15 Ala Phe Phe Leu Ser Thr Glu Lys Lys Asn Leu Glu Thr Lys
 160 165 170

<210> 353
 <211> 88
 20 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 25 <222> -44...-1

<400> 353
 Met Ala Ala Glu Gly Trp Ile Trp Arg Trp Gly Trp Gly Arg Arg Cys
 -40 -35 -30
 30 Leu Gly Arg Pro Gly Leu Leu Gly Pro Gly Pro Gly Pro Thr Thr Pro
 -25 -20 -15
 Leu Phe Leu Leu Leu Leu Leu Gly Ser Val Thr Ala Asp Ile Thr Asp
 -10 -5 1
 Gly Asn Ile Glu His Leu Lys Arg Glu His Ser Leu Ile Lys Pro Tyr
 35 5 10 15 20
 Gln Gly Val Gly Ser Ser Ser Pro Ser Gly Thr Ser Arg Ala Ala Leu
 25 30 35
 Cys Ser Arg Ala Ser Thr Tyr Val
 40

40
 <210> 354
 <211> 151
 <212> PRT
 <213> Homo sapiens

45
 <220>
 <221> SIGNAL
 <222> -32...-1

50 <400> 354
 Met Asp Ser Ala Ser Asn Pro Thr Asn Leu Val Ser Thr Ser Gln Arg
 -30 -25 -20
 His Arg Pro Leu Leu Ser Ser Cys Gly Leu Pro Pro Ser Thr Ala Ser
 -15 -10 -5
 55 Ala Val Arg Arg Leu Cys Ser Arg Gly Val Leu Lys Gly Ser Asn Glu
 1 5 10 15
 Arg Arg Asp Met Glu Ser Phe Trp Lys Leu Asn Arg Ser Pro Gly Ser
 20 25 30
 Asp Arg Tyr Leu Glu Ser Arg Asp Ala Ser Arg Leu Ser Gly Arg Asp
 60 35 40 45
 Pro Ser Ser Trp Thr Val Glu Asp Val Met Gln Phe Val Arg Glu Ala
 50 55 60
 Asp Pro Gln Leu Gly Pro His Ala Asp Leu Phe Arg Lys His Glu Ile

305

His Gly Leu Ser Ser Leu Cys Asn Leu Gly Cys Val Leu Ser Asn Gly
 140 145 150
 Leu Cys Leu Ala Gly Leu Ala Leu Glu Ile Arg Ser Leu
 155 160 165

5

<210> 357
 <211> 183
 <212> PRT
 <213> Homo sapiens

10

<220>
 <221> SIGNAL
 <222> -47...-1

15

<400> 357
 Met Thr Glu Cys Thr Ser Leu Gln Phe Val Ser Pro Phe Ala Phe Glu
 -45 -40 -35
 Ala Met Gln Lys Val Asp Val Val Cys Leu Ala Ser Leu Ser Asp Pro
 -30 -25 -20

20

Glu Leu Arg Leu Leu Leu Pro Cys Leu Val Arg Met Ala Leu Cys Ala
 -15 -10 -5 1
 Pro Ala Asp Gln Ser Gln Ser Trp Ala Gln Asp Lys Lys Leu Ile Leu
 5 10 15

25

Arg Leu Leu Ser Gly Val Glu Ala Val Asn Ser Ile Val Ala Leu Leu
 20 25 30
 Ser Val Asp Phe His Ala Leu Glu Gln Asp Ala Ser Lys Glu Gln Gln
 35 40 45
 Leu Arg Pro Ser Leu Ala Leu Leu Pro Arg Leu Glu Cys Gly Gly Val
 50 55 60 65

30

Ile Ser Ala His Cys Asn Leu His Leu Leu Gly Ser Ser Asp Ser Ser
 70 75 80
 Ala Ser Val Ser Arg Val Asp Gly Thr Thr Gly Thr Arg His His Ala
 85 90 95

35

Arg Leu Phe Cys Ile Ile Ser Arg Asp Glu Val Ser Pro Tyr Trp Pro
 100 105 110
 Gly Trp Ser Arg Thr Pro Asn Leu Val Ile His Leu Pro Gln Pro Pro
 115 120 125
 Lys Val Leu Gly Leu Pro Ala
 130 135

40

<210> 358
 <211> 102
 <212> PRT
 <213> Homo sapiens

45

<220>
 <221> SIGNAL
 <222> -14...-1

50

<400> 358
 Met Phe Leu Thr Ala Leu Leu Trp Arg Gly Arg Ile Pro Gly Arg Gln
 -10 -5 1
 Trp Ile Gly Lys His Arg Arg Pro Arg Phe Val Ser Leu Arg Ala Lys
 5 10 15

55

Gln Asn Met Ile Arg Arg Leu Glu Ile Glu Ala Glu Asn His Tyr Trp
 20 25 30
 Leu Ser Met Pro Tyr Met Thr Arg Glu Gln Glu Arg Gly His Ala Ala
 35 40 45 50
 Val Arg Arg Arg Glu Ala Phe Glu Ala Ile Lys Ala Ala Ala Thr Ser
 55 60 65

60

Lys Phe Pro Pro His Arg Phe Ile Ala Asp Gln Leu Asp His Leu Asn
 70 75 80
 Val Thr Lys Lys Trp Ser

```

<210> 359
<211> 244
5 <212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
10 <222> -29...-1

<400> 359
Met Glu Leu Thr Ile Phe Ile Leu Arg Leu Ala Ile Tyr Ile Leu Thr
-25 -20 -15
15 Phe Pro Leu Tyr Leu Leu Asn Phe Leu Gly Leu Trp Ser Trp Ile Cys
-10 -5 1
Lys Lys Trp Phe Pro Tyr Phe Leu Val Arg Phe Thr Val Ile Tyr Asn
5 10 15
Glu Gln Met Ala Ser Lys Lys Arg Glu Leu Phe Ser Asn Leu Gln Glu
20 20 25 30 35
Phe Ala Gly Pro Ser Gly Lys Leu Ser Leu Leu Glu Val Gly Cys Gly
40 45 50
Thr Gly Ala Asn Phe Lys Phe Tyr Pro Pro Gly Cys Arg Val Thr Cys
55 60 65
25 Ile Asp Pro Asn Pro Asn Phe Glu Lys Phe Leu Ile Lys Ser Ile Ala
70 75 80
Glu Asn Arg His Leu Gln Phe Glu Arg Phe Val Val Ala Ala Gly Glu
85 90 95
Asn Met His Gln Val Ala Asp Gly Ser Val Asp Val Val Val Cys Thr
30 100 105 110 115
Leu Val Leu Cys Ser Val Lys Asn Gln Glu Arg Ile Leu Arg Glu Val
120 125 130
Cys Arg Val Leu Arg Pro Gly Gly Ala Phe Tyr Phe Met Glu His Val
135 140 145
35 Ala Ala Glu Cys Ser Thr Trp Asn Tyr Phe Trp Gln Gln Val Leu Asp
150 155 160
Pro Ala Trp His Leu Leu Phe Asp Gly Cys Asn Leu Thr Arg Glu Ser
165 170 175
40 Trp Lys Ala Leu Glu Arg Ala Ser Phe Ser Lys Leu Lys Leu Gln His
180 185 190 195
Ile Gln Ala Pro Leu Ser Trp Glu Leu Val Arg Pro His Ile Tyr Gly
200 205 210
Tyr Ala Val Lys
215

45
<210> 360
<211> 177
<212> PRT
<213> Homo sapiens

50
<220>
<221> SIGNAL
<222> -23...-1

55 <400> 360
Met Ser Asn Gln Arg Leu Pro Leu Ile Phe Ser Leu Leu Phe Ile Cys
-20 -15 -10
Phe Phe Gly Glu Ser Phe Cys Ile Cys Asp Gly Thr Val Trp Thr Lys
-5 1 5
60 Val Gly Trp Glu Ile Leu Pro Glu Glu Val His Tyr Trp Lys Val Lys
10 15 20 25
Gly Ser Pro Ser His Cys Leu Pro Tyr Leu Leu Asp Lys Leu Cys Cys
30 35 40

```

Asp Phe Ala Asn Met Asp Ile Phe Gln Gly Cys Leu Tyr Leu Ile Tyr
 45 50 55
 Asn Leu Leu Gln Ala Val Phe Phe Val Leu Phe Val Leu Ser Val His
 60 65 70
 5 Tyr Leu Trp Lys Lys Trp Lys Lys His Gln Lys Lys Leu Lys Lys Gln
 75 80 85
 Ala Ser Leu Glu Lys Pro Gly Asn Asp Leu Glu Ser Pro Leu Ile Asn
 90 95 100 105
 10 Asn Ile Asp Gln Thr Leu His Arg Val Ala Thr Thr Ala Ser Val Ile
 110 115 120
 Tyr Lys Ile Trp Glu His Arg Ser His His Pro Ser Ser Lys Lys Ile
 125 130 135
 Lys His Cys Lys Leu Lys Lys Lys Ser Lys Glu Glu Gly Ala Arg Arg
 140 145 150
 15 Tyr

<210> 361

<211> 158

<212> PRT

20 <213> Homo sapiens

<220>

<221> SIGNAL

<222> -21...-1

25

<400> 361

Met Ala Leu Cys Ala Leu Thr Arg Ala Leu Pro Ser Leu Asn Leu Ala
 -20 -15 -10
 30 Pro Pro Thr Val Ala Ala Pro Ala Pro Ser Leu Phe Pro Ala Ala Gln
 -5 1 5 10
 Met Met Asn Asn Gly Leu Leu Gln Gln Pro Ser Ala Leu Met Leu Leu
 15 20 25
 Pro Cys Arg Pro Val Leu Thr Ser Val Ala Leu Asn Ala Asn Phe Val
 30 35 40
 35 Ser Trp Lys Ser Arg Thr Lys Tyr Thr Ile Thr Pro Val Lys Met Arg
 45 50 55
 Lys Ser Gly Gly Arg Asp His Thr Gly Ala Gly Asn Val Arg Arg Thr
 60 65 70 75
 Val Gly Arg Val Ser Asn Val Asp His Asn Lys Arg Val Ile Gly Lys
 80 85 90
 40 Ala Gly Arg Asn Arg Trp Leu Gly Lys Arg Pro Asn Ser Gly Arg Trp
 95 100 105
 His Arg Lys Gly Gly Trp Ala Gly Arg Lys Ile Arg Pro Leu Pro Pro
 110 115 120
 45 Met Lys Ser Tyr Val Lys Leu Pro Ser Ala Ser Ala Gln Ser
 125 130 135

<210> 362

<211> 186

50 <212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

55 <222> -19...-1

<400> 362

Met Ala Thr Ala Ser Pro Ser Val Phe Leu Leu Met Val Asn Gly Gln
 -15 -10 -5
 60 Val Glu Ser Ala Gln Phe Pro Glu Tyr Asp Asp Leu Tyr Cys Lys Tyr
 1 5 10
 Cys Phe Val Tyr Gly Gln Asp Trp Ala Pro Thr Ala Gly Leu Glu Glu
 15 20 25

Gly Ile Ser Gln Ile Thr Ser Lys Ser Gln Asp Val Arg Gln Ala Leu
 30 35 40 45
 Val Trp Asn Phe Pro Ile Asp Val Thr Phe Lys Ser Thr Asn Pro Tyr
 50 55 60
 5 Gly Trp Pro Gln Ile Val Leu Ser Val Tyr Gly Pro Asp Val Phe Gly
 65 70 75
 Asn Asp Val Val Arg Gly Tyr Gly Ala Val His Val Pro Phe Ser Pro
 80 85 90
 Gly Arg His Lys Arg Thr Ile Pro Met Phe Val Pro Glu Ser Thr Ser
 10 95 100 105
 Lys Leu Gln Lys Phe Thr Ser Trp Phe Met Gly Arg Arg Pro Glu Tyr
 110 115 120 125
 Thr Asp Pro Lys Val Val Ala Gln Gly Glu Gly Arg Glu Ala Ile Thr
 130 135 140
 15 Ala Pro Arg Lys Ala Val Phe Ser Val His Gly Leu Thr Ser Pro Arg
 145 150 155
 Ala Leu Ala Leu Val His Ile Lys Gly Thr
 160 165

20 <210> 363
 <211> 150
 <212> PRT
 <213> Homo sapiens

25 <220>
 <221> SIGNAL
 <222> -47..-1

<400> 363
 30 Met Gly Asp Arg Val Lys Gly Ser Lys Ser Arg Ala Phe Val Ser Pro
 -45 -40 -35
 Trp Pro His Thr Pro Met Ala Ser Gly Leu Arg Asp Pro Trp Leu Gln
 -30 -25 -20
 Pro Thr Ala Leu Gly Leu Ala Leu Cys Ser Thr Lys Ala Leu Ser Val
 35 -15 -10 -5 1
 Gly Ser Ala Pro Leu Pro Pro Arg Asn Ser Asn Thr Met Ala Ala Ala
 5 10 15
 Ala Leu Ala Ala Pro Ser Leu Gly Phe Asp Gly Val Ile Gly Val Leu
 20 25 30
 40 Val Ala Asp Thr Ser Leu Thr Asp Met His Val Val Asp Val Glu Leu
 35 40 45
 Ser Gly Pro Arg Gly Pro Thr Gly Arg Ser Phe Ala Val His Thr Arg
 50 55 60 65
 Arg Glu Asn Pro Ala Glu Pro Gly Ala Val Thr Gly Ser Ala Thr Val
 45 70 75 80
 Thr Ala Phe Trp Arg Ser Leu Leu Ala Cys Cys Gln Leu Pro Ser Arg
 85 90 95
 Pro Gly Ile His Leu Cys
 100

50
 <210> 364
 <211> 95
 <212> PRT
 <213> Homo sapiens

55 <220>
 <221> SIGNAL
 <222> -45..-1

60 <400> 364
 Met Leu His His Val Ile Thr Ala Gly Pro Val Leu Leu Leu His Leu
 -45 -40 -35 -30
 Pro Arg Pro Asp Thr Ser Thr Arg Leu Leu Thr Ser Val Ser Ala

-25 -20 -15
 Phe Ile Leu Leu Leu Leu Ser Gly Pro Ala Glu Met Ser Ala Ser
 -10 -5 1
 Gln Glu Ser Phe Pro Gly Ser Leu Gln Gln Glu Ile Ala Ser Leu Ile
 5 5 10 15
 Thr Val Ala Leu Gly Ser Leu Ile Ser Leu Ser Cys Ser Thr Leu Leu
 20 25 30 35
 Tyr Phe Ser Cys Glu Leu Lys Ile Pro Cys Glu Asp Val Asn Leu
 40 45 50
 10
 <210> 365
 <211> 94
 <212> PRT
 <213> Homo sapiens
 15
 <220>
 <221> SIGNAL
 <222> -26...-1
 20 <400> 365
 Met Ala Ala Ile Glu Ile Glu Val Lys Pro Asn Gln Gly Phe Cys Gly
 -25 -20 -15
 Ser Ala Cys Leu Leu Ala Val Ile Arg Ala Phe Phe Phe Lys Lys Asn
 -10 -5 1 5
 25 Ala Cys Leu Leu Arg Glu Ile Leu Gln Ser Lys Leu Gly Gly Met Gly
 10 15 20
 Pro Val Val Phe Ser Tyr Arg Gly Leu Pro Leu Trp Leu Phe Ala Trp
 25 30 35
 Leu Phe Pro Arg Cys Thr Val Pro Leu Thr Phe Gly Phe Glu Asn Met
 30 40 45 50
 Arg Gly Leu Gly Val Val Ala Tyr Ala Cys Asn Pro Ser Thr
 55 60 65
 <210> 366
 35 <211> 140
 <212> PRT
 <213> Homo sapiens
 <220>
 40 <221> SIGNAL
 <222> -40...-1
 <400> 366
 45 Met Thr Ser Met Thr Gln Ser Leu Arg Glu Val Ile Lys Ala Met Thr
 -40 -35 -30 -25
 Lys Ala Arg Asn Phe Glu Arg Val Leu Gly Lys Ile Thr Leu Val Ser
 -20 -15 -10
 Ala Ala Pro Gly Lys Val Ile Cys Glu Met Lys Val Glu Glu Glu His
 -5 1 5
 50 Thr Asn Ala Ile Gly Thr Leu His Gly Gly Leu Thr Ala Thr Leu Val
 10 15 20
 Asp Asn Ile Ser Thr Met Ala Leu Leu Cys Thr Glu Arg Gly Ala Pro
 25 30 35 40
 Gly Val Ser Val Asp Met Asn Ile Thr Tyr Met Ser Pro Ala Lys Leu
 55 45 50 55
 Gly Glu Asp Ile Val Ile Thr Ala His Val Leu Lys Gln Gly Lys Thr
 60 65 70
 Leu Ala Phe Thr Ser Val Asp Leu Thr Asn Lys Ala Thr Gly Lys Leu
 75 80 85
 60 Ile Ala Gln Gly Arg His Thr Lys His Leu Gly Asn
 90 95 100
 <210> 367

<211> 39
 <212> PRT
 <213> Homo sapiens

5 <220>
 <221> SIGNAL
 <222> -35...-1

<400> 367
 10 Met Asp Pro Gly Trp Pro His Phe Lys Leu Thr His Ser Arg Cys Met
 -35 -30 -25 -20
 Ala Val Leu Phe Leu Gly Thr Leu Pro Leu Cys Pro Val Thr Ser Pro
 -15 -10 -5
 Val Trp Gly Trp Ser Pro Gly
 15 1

<210> 368
 <211> 78
 <212> PRT

20 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -41...-1

25 <400> 368
 Met Ser Ala Ser Val Val Ser Val Ile Ser Arg Phe Leu Glu Glu Tyr
 -40 -35 -30
 30 Leu Ser Ser Thr Pro Gln Arg Leu Lys Leu Leu Asp Ala Tyr Leu Leu
 -25 -20 -15 -10
 Tyr Ile Leu Leu Thr Gly Ala Leu Gln Phe Gly Tyr Cys Leu Leu Val
 -5 1 5
 Gly Thr Phe Pro Phe Asn Ser Phe Leu Ser Gly Phe Ile Ser Cys Val
 10 15 20
 35 Gly Ser Phe Ile Leu Ala Gly Ser Leu Phe Glu Phe Pro Gly
 25 30 35

<210> 369
 <211> 83
 40 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 45 <222> -40...-1

<400> 369
 Met Gly Leu Thr Ser Thr Trp Arg Tyr Gly Arg Gly Pro Gly Ile Gly
 -40 -35 -30 -25
 50 Thr Val Thr Met Val Ser Trp Gly Arg Phe Ile Cys Leu Val Val Val
 -20 -15 -10
 Thr Met Ala Thr Leu Ser Leu Ala Arg Pro Ser Phe Ser Leu Val Glu
 -5 1 5
 Asp Thr Thr Leu Glu Pro Glu Asp Ala Ile Ser Ser Gly Asp Asp Glu
 55 10 15 20
 Asp Asp Thr Asp Gly Ala Glu Asp Phe Val Ser Glu Asn Ser Asn Asn
 25 30 35 40
 Lys Ser Lys

60 <210> 370
 <211> 92
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -15...-1
 5
 <400> 370
 Met Ala Val Leu Ala Gly Ser Leu Leu Gly Pro Thr Ser Arg Ser Ala
 -15 -10 -5 1
 10 Ala Leu Leu Gly Gly Arg Trp Leu Gln Pro Arg Ala Trp Leu Gly Phe
 5 10 15
 Pro Asp Ala Trp Gly Leu Pro Thr Pro Gln Gln Ala Arg Gly Lys Ala
 20 25 30
 Arg Gly Asn Glu Tyr Gln Pro Ser Asn Ile Lys Arg Lys Asn Lys His
 35 40 45
 15 Gly Trp Val Arg Arg Leu Ser Thr Pro Ala Gly Val Gln Val Ile Leu
 50 55 60 65
 Arg Arg Met Leu Lys Gly Arg Lys Ser Leu Ser His
 70 75
 20 <210> 371
 <211> 279
 <212> PRT
 <213> Homo sapiens
 25 <220>
 <221> SIGNAL
 <222> -42...-1
 <400> 371
 30 Met Ala Ala Pro Val Arg Arg Thr Leu Leu Gly Val Ala Gly Gly Trp
 -40 -35 -30
 Arg Arg Phe Glu Arg Leu Trp Ala Gly Ser Leu Ser Ser Arg Ser Leu
 -25 -20 -15
 35 Ala Leu Ala Ala Ala Pro Ser Ser Asn Gly Ser Pro Trp Arg Leu Leu
 -10 -5 1 5
 Gly Ala Leu Cys Leu Gln Arg Pro Pro Val Val Ser Lys Pro Leu Thr
 10 15 20
 Pro Leu Gln Glu Glu Met Ala Ser Leu Leu Gln Gln Ile Glu Ile Glu
 25 30 35
 40 Arg Ser Leu Tyr Ser Asp His Glu Leu Arg Ala Leu Asp Glu Asn Gln
 40 45 50
 Arg Leu Ala Lys Lys Lys Ala Asp Leu His Asp Glu Glu Asp Glu Gln
 55 60 65 70
 45 Asp Ile Leu Leu Ala Gln Asp Leu Glu Asp Met Trp Glu Gln Lys Phe
 75 80 85
 Leu Gln Phe Lys Leu Gly Ala Arg Ile Thr Glu Ala Asp Glu Lys Asn
 90 95 100
 Asp Arg Thr Ser Leu Asn Arg Asn Leu Asp Arg Asn Leu Val Leu Leu
 105 110 115
 50 Val Arg Glu Lys Phe Gly Asp Gln Asp Val Trp Ile Leu Pro Gln Ala
 120 125 130
 Glu Trp Gln Pro Gly Glu Thr Leu Arg Gly Thr Ala Glu Arg Thr Leu
 135 140 145 150
 Ala Thr Leu Ser Glu Asn Asn Met Glu Ala Lys Phe Leu Gly Asn Ala
 155 160 165
 55 Pro Cys Gly His Tyr Thr Phe Lys Phe Pro Gln Ala Met Arg Thr Glu
 170 175 180
 Ser Asn Leu Gly Ala Lys Val Phe Phe Phe Lys Ala Leu Leu Leu Thr
 185 190 195
 60 Gly Asp Phe Ser Gln Ala Gly Asn Lys Gly His His Val Trp Val Ile
 200 205 210
 Lys Asp Glu Leu Gly Asp Tyr Leu Lys Pro Lys Tyr Leu Ala Gln Val
 215 220 225 230

Arg Arg Phe Val Ser Asp Leu
235

<210> 372
5 <211> 184
<212> PRT
<213> Homo sapiens

<220>
10 <221> SIGNAL
<222> -31...-1

<400> 372
Met Ala Cys Thr Thr Thr Ala Pro Ala Gln Glu His Met Leu Leu Thr
15 -30 -25 -20
Pro Leu Thr Ala Leu Met Val Gly Ala Ala Ser Leu Leu Glu Gly Arg
-15 -10 -5 1
Pro Gln Ile Ser Ala Pro Tyr Ser Arg Ala Ala Cys Cys Ser Pro Gly
5 10 15
20 Ala Leu Gly Cys Pro Ala Ala Arg Val Gly Ile Leu Asp Leu Met Tyr
20 25 30
Ser Trp Val Ala Arg Lys Val Leu Arg Cys Ser Asn Thr Gly Leu Gln
35 40 45
Gly Leu His Cys Ala Pro Ala Tyr Ala Ala Gln Leu Gly Met Asp Pro
25 50 55 60 65
Gly Arg Gly Gln Arg Ala Gly Gly Pro Val Glu Gln Thr Tyr Phe Ser
70 75 80
Pro Met Gly Lys Leu Pro Thr Leu Ser Trp Leu Glu Gly Cys Thr Ala
85 90 95
30 Val Met Thr Leu Ala Ser Ala Trp Leu Leu Gly Ser Pro Arg Glu Thr
100 105 110
Tyr Asn His Glu Lys Val Lys Glu Lys Gln Cys Pro Phe Ser Ser Met
115 120 125
Val Leu Gly Glu Tyr Gly Phe Leu Pro Thr Val Asp His Leu Ser Thr
35 130 135 140 145
Leu Gly Cys Asn Met Arg Glu Leu
150

<210> 373
40 <211> 101
<212> PRT
<213> Homo sapiens

<220>
45 <221> SIGNAL
<222> -42...-1

<400> 373
Met Ala His Val Ala Glu Lys Asp Gly Leu Asp Trp Ala Ser Gly Cys
50 -40 -35 -30
Ile Pro Gly Leu Gln Thr Gly Ile Cys Leu Phe Gly Ser Gln Leu Cys
-25 -20 -15
Phe His Leu Ser Trp Leu Tyr Ser Trp Ala Ser Gln Cys Gly Pro Thr
-10 -5 1 5
55 Ala Pro Val Ile Asp Lys Lys Ser Ser Pro Leu Leu Thr Glu Leu Leu
10 15 20
Asp Leu Val Leu Ile Gly Pro Asp Glu Glu Gly Ile Gln Pro Gln Val
25 30 35
Ile Ile Val Ala Arg Lys Met Glu Tyr Thr Lys Trp Thr Gly Leu Ala
60 40 45 50
Cys Thr His Arg Asp
55

<210> 374
 <211> 85
 <212> PRT
 <213> Homo sapiens
 5
 <220>
 <221> SIGNAL
 <222> -20...-1
 10 <400> 374
 Met Gly Pro Asn Thr Lys Asn Leu Leu Leu Val Thr Leu Val Ala Ser
 -20 -15 -10 -5
 Thr Val Pro Gly Asn Ser Leu Gly Gln Asp Phe Thr Phe Ala His Leu
 1 5 10
 15 Glu Arg Ser Cys Thr Arg Glu Asn Arg Ser Pro Gly Glu Val Phe Gln
 15 20 25
 Gln Pro Cys Lys Ser Gly Gly Gly Gly Val Gly Glu Pro Asn Ala Gln
 30 35 40
 Gly Gln Leu Leu Ser Gln His Pro Leu Pro Ala Phe Ile Asn Cys Ser
 20 45 50 55 60
 His Gly Gln Ala Phe
 65
 <210> 375
 25 <211> 90
 <212> PRT
 <213> Homo sapiens
 <220>
 30 <221> SIGNAL
 <222> -28...-1
 <400> 375
 Met Ala Phe Pro Gly Gln Ser Asp Thr Lys Met Gln Trp Pro Glu Val
 35 -25 -20 -15
 Pro Ala Leu Pro Leu Leu Ser Ser Leu Cys Met Ala Met Val Arg Lys
 -10 -5 1
 Ser Ser Ala Leu Gly Lys Glu Val Gly Arg Arg Val Lys Glu Met Val
 5 10 15 20
 40 Met Leu Val Ala Pro Phe Arg Gln Ser Ser Ser Leu Ser Arg Thr Phe
 25 30 35
 Ser Ser Arg Lys Val Val Lys Ala His Ala Ser Leu His Gly Ala Arg
 40 45 50
 Leu Ser Pro Leu Ser Arg Asn Ile Arg Gly
 45 55 60
 <210> 376
 <211> 89
 <212> PRT
 50 <213> Homo sapiens
 <220>
 <221> SIGNAL
 <222> -33...-1
 55
 <220>
 <221> UNSURE
 <222> 47
 <223> Xaa = Ala,Pro,Ser,Thr
 60
 <400> 376
 Met Ala Gln Pro Ala Ala Pro Ser Leu Thr Arg Pro Phe Leu Ala Glu
 -30 -25 -20

Ala Pro Thr Ala Leu Val Pro His Ser Pro Leu Pro Gly Ala Leu Ser
 -15 -10 -5
 Ser Ala Pro Gly Pro Lys Gln Pro Pro Thr Ala Ser Thr Gly Pro Glu
 1 5 10 15
 5 Leu Leu Leu Leu Pro Leu Ser Ser Phe Met Pro Cys Gly Ala Ala Ala
 20 25 30
 Pro Ala Arg Val Ser Ser Gln Arg Ala Thr Pro Arg Asp Lys Pro Xaa
 35 40 45
 Gly Pro Leu Ile Pro Gly Gln Cys Pro
 10 50 55

<210> 377

<211> 132

<212> PRT

15 <213> Homo sapiens

<220>

<221> SIGNAL

<222> -15...-1

20

<400> 377

Met Asn Arg Val Leu Cys Ala Pro Ala Ala Gly Ala Val Arg Ala Leu
 -15 -10 -5 1
 Arg Leu Ile Gly Trp Ala Ser Arg Ser Leu His Pro Leu Pro Gly Ser
 25 5 10 15
 Arg Asp Arg Ala His Pro Ala Ala Glu Glu Glu Asp Asp Pro Asp Arg
 20 25 30
 Pro Ile Glu Phe Ser Ser Ser Lys Ala Asn Pro His Arg Trp Ser Val
 35 40 45
 30 Gly His Thr Met Gly Lys Gly His Gln Arg Pro Trp Trp Lys Val Leu
 50 55 60 65
 Pro Leu Ser Cys Phe Leu Val Ala Leu Ile Ile Trp Cys Tyr Leu Arg
 70 75 80
 Glu Glu Ser Glu Ala Asp Gln Trp Leu Arg Gln Val Trp Gly Glu Val
 35 85 90 95
 Pro Glu Pro Ser Asp Arg Ser Glu Glu Pro Glu Thr Pro Ala Ala Tyr
 100 105 110
 Arg Ala Arg Thr
 115

40

<210> 378

<211> 102

<212> PRT

<213> Homo sapiens

45

<220>

<221> SIGNAL

<222> -14...-1

50

<220>

<221> UNSURE

<222> 50

<223> Xaa = Ala,Gly

55

<220>

<221> UNSURE

<222> 51

<223> Xaa = Leu,Met,Val

60

<400> 378

Met Phe Leu Thr Ala Leu Leu Trp Arg Gly Arg Ile Pro Gly Arg Gln
 -10 -5 1
 Trp Ile Gly Lys His Arg Arg Pro Arg Phe Val Ser Leu Arg Ala Lys

	<400> 379															
	Met	Gly	Ile	Lys	Thr	Ala	Leu	Pro	Ala	Ala	Glu	Leu	Gly	Leu	Tyr	Ser
					-20					-15					-10	
25	Leu	Val	Leu	Ser	Gly	Ala	Leu	Ala	Tyr	Ala	Gly	Arg	Gly	Leu	Leu	Glu
				-5					1				5			
	Ala	Ser	Gln	Asp	Gly	Ala	His	Arg	Lys	Ala	Phe	Arg	Glu	Ser	Val	Arg
	10						15					20				
	Pro	Gly	Trp	Glu	Tyr	Ile	Gly	Arg	Lys	Met	Asp	Val	Ala	Asp	Phe	Glu
30	25					30					35					40
	Trp	Val	Met	Trp	Phe	Thr	Ser	Phe	Arg	Asn	Val	Ile	Ile	Phe	Ala	Leu
					45					50					55	
	Ser	Gly	His	Val	Leu	Phe	Ala	Lys	Leu	Cys	Thr	Met	Val	Ala	Pro	Lys
				60					65					70		
35	Leu	Arg	Ser	Trp	Met	Tyr	Ala	Val	Tyr	Gly	Ala	Leu	Ala	Val	Met	Gly
			75					80					85			
	Thr	Met	Gly	Pro	Trp	Tyr	Leu	Leu	Leu	Leu	Leu	Gly	His	Cys	Val	Gly
	90						95					100				
	Leu	Tyr	Val	Ala	Ser	Leu	Leu	Gly	Gln	Pro	Trp	Leu	Cys	Leu	Gly	Leu
40	105					110					115					120
	Gly	Leu	Ala	Ser	Leu	Ala	Ser	Phe	Lys	Met	Asp	Pro	Leu	Ile	Ser	Trp
					125					130					135	
	Gln	Ser	Gly	Phe	Val	Thr	Gly	Thr	Phe	Asp	Leu	Gln	Glu	Val	Leu	Phe
				140					145					150		
45	His	Gly	Gly	Ser	Ser	Phe	Thr	Val	Leu	Arg	Cys	Thr	Ser	Phe	Ala	Leu
			155					160					165			
	Glu	Ser	Cys	Ala	His	Pro	Asp	Arg	His	Tyr	Ser	Leu	Ala	Asp	Leu	Leu
	170						175					180				
	Lys	Tyr	Ser	Phe	Tyr	Leu	Pro	Phe	Phe	Phe	Phe	Gly	Pro	Ile	Met	Thr
50	185					190					195					200
	Phe	Asp	Arg	Phe	His	Ala	Gln	Val	Ser	Gln	Val	Glu	Pro	Val	Arg	Arg
					205					210					215	
	Glu	Gly	Glu	Leu	Trp	His	Ile	Arg	Ala	Gln	Ala	Gly	Leu	Ser	Val	Val
				220					225					230		
55	Ala	Ile	Met	Ala	Val	Asp	Ile	Phe	Phe	His	Phe	Phe	Tyr	Ile	Leu	Thr
			235					240					245			
	Ile	Pro	Ser	Asp	Leu	Lys	Phe	Ala	Asn	Arg	Leu	Pro	Asp	Ile	Ala	Leu
	250						255					260				
	Ala	Gly	Leu	Ala	Tyr	Ser	Asn	Leu	Val	Tyr	Asp	Trp	Val	Lys	Ala	Ala
60	265					270					275					280
	Val	Leu	Phe	Gly	Val	Val	Asn	Thr	Val	Ala	Cys	Leu	Asp	His	Leu	Asp
					285											

```

      300      305      310
Thr His Phe Asp Arg Gly Ile Asn Asp Trp Leu Cys Lys Tyr Val Tyr
      315      320      325
5  Asn His Ile Gly Gly Glu His Ser Ala Val Ile Pro Glu Leu Ala Ala
      330      335      340
Thr Val Ala Thr Phe Ala Ile Thr Thr Leu Trp Leu Gly Pro Cys Asp
345      350      355      360
Ile Val Tyr Leu Trp Ser Phe Leu Asn Cys Phe Gly Leu Asn Phe Glu
      365      370      375
10 Leu Trp Met Gln Lys Leu Ala Glu Trp Gly Pro Leu Ala Arg Ile Glu
      380      385      390
Ala Ser Leu Ser Val Gln Met Ser Arg Arg Val Arg Ala Leu Phe Gly
      395      400      405
15 Ala Met Asn Phe Trp Ala Ile Ile Met Tyr Asn Leu Val Ser Leu Asn
      410      415      420
Ser Leu Lys Phe Thr Glu Leu Val Ala Arg Arg Leu Leu Leu Thr Gly
425      430      435      440
Phe Pro Gln Thr Thr Leu Ser Ile Leu Phe Val Thr Tyr Cys Gly Val
      445      450      455
20 Gln Leu Val Lys Glu Arg Glu Arg Thr Leu Ala Leu Glu Glu Glu Gln
      460      465      470
Lys Gln Asp Lys Glu Lys Pro Glu
      475      480

```

25 <210> 380
 <211> 152
 <212> PRT
 <213> Homo sapiens

30 <220>
 <221> SIGNAL
 <222> -26...-1

```

      <400> 380
35 Met Val Thr Phe Pro Asp Val Pro Leu Gly Ile Phe Leu Phe Cys Val
      -25      -20      -15
Cys Val Ile Ala Ile Gly Val Val Gln Ala Leu Ile Val Gly Tyr Ala
-10      -5      1      5
Phe His Phe Pro His Leu Leu Ser Pro Gln Ile Gln Arg Ser Ala His
40      10      15      20
Arg Ala Leu Tyr Arg Arg His Val Leu Gly Ile Val Leu Gln Gly Pro
      25      30      35
Ala Leu Cys Phe Ala Ala Ala Ile Phe Ser Leu Phe Phe Val Pro Leu
40      45      50
45 Ser Tyr Leu Leu Met Val Thr Val Ile Leu Leu Pro Tyr Val Ser Lys
55      60      65      70
Val Thr Gly Trp Cys Arg Asp Arg Leu Leu Gly His Arg Glu Pro Ser
      75      80      85
Ala His Pro Val Glu Val Phe Ser Phe Asp Leu His Glu Pro Leu Ser
50      90      95      100
Lys Glu Arg Val Glu Ala Phe Ser Asp Gly Val Tyr Ala Ile Val Ala
      105      110      115
Thr Leu Leu Ile Leu Asp Ile Trp
      120      125

```

55 <210> 381
 <211> 51
 <212> PRT
 <213> Homo sapiens

60 <220>
 <221> SIGNAL
 <222> -26...-1

<400> 381
 Met Glu Met Leu Phe Asp Glu Arg Ala Pro Leu Leu Phe Ile Leu Phe
 -25 -20 -15
 5 Lys Phe Ser Leu Cys Pro Tyr Ala Ala Ala Leu Ser Lys Pro Ile Phe
 -10 -5 1 5
 Gly Ser Val Ala Cys Met Thr Lys Glu Ile Leu Ala Arg His Gly Gly
 10 15 20
 Ser Arg Leu
 10 25

 <210> 382
 <211> 72
 <212> PRT
 15 <213> Homo sapiens

 <220>
 <221> SIGNAL
 <222> -23...-1
 20
 <400> 382
 Met Leu Arg Pro Ala Leu Pro Trp Leu Tyr Leu Gly Leu Cys Ser Leu
 -20 -15 -10
 Leu Val Gly Glu Ala Glu Ala Pro Ser Pro Val Asp Pro Leu Glu Arg
 25 -5 1 5
 Ser Arg Pro Tyr Ala Val Leu Arg Gly Gln Asn Leu Val Leu Met Gly
 10 15 20 25
 Thr Ile Phe Ser Ile Leu Leu Val Thr Val Ile Leu Met Ala Phe Cys
 30 35 40
 30 Val Tyr Lys Pro Ile Arg Arg Arg
 45

 <210> 383
 <211> 95
 35 <212> PRT
 <213> Homo sapiens

 <220>
 <221> SIGNAL
 40 <222> -48...-1

 <400> 383
 Met Ala Ser Ser His Trp Asn Glu Thr Thr Thr Ser Val Tyr Gln Tyr
 -45 -40 -35
 45 Leu Gly Phe Gln Val Gln Lys Ile Tyr Pro Phe His Asp Asn Trp Asn
 -30 -25 -20
 Thr Ala Cys Phe Val Ile Leu Leu Leu Phe Ile Phe Thr Val Val Ser
 -15 -10 -5
 Leu Val Val Leu Ala Phe Leu Tyr Glu Val Leu Asp Cys Cys Cys Cys
 50 1 5 10 15
 Val Lys Asn Lys Thr Val Lys Asp Leu Lys Ser Glu Pro Asn Pro Leu
 20 25 30
 Arg Ser Met Met Asp Asn Ile Arg Lys Arg Glu Thr Glu Val Val
 35 40 45
 55
 <210> 384
 <211> 150
 <212> PRT
 <213> Homo sapiens
 60
 <220>
 <221> SIGNAL
 <222> -20...-1

<400> 384
 Met Ala Arg His Gly Leu Pro Leu Leu Pro Leu Leu Ser Leu Leu Val
 -20 -15 -10 -5
 5 Gly Ala Trp Leu Lys Leu Gly Asn Gly Gln Ala Thr Ser Met Val Gln
 1 5 10
 Leu Gln Gly Gly Arg Phe Leu Met Gly Thr Asn Ser Pro Asp Ser Arg
 15 20 25
 Asp Gly Glu Gly Pro Val Arg Glu Ala Thr Val Lys Pro Phe Ala Ile
 30 35 40
 10 Asp Ile Phe Pro Val Thr Asn Lys Asp Phe Arg Asp Phe Val Arg Glu
 45 50 55 60
 Lys Lys Tyr Arg Thr Glu Ala Glu Met Phe Gly Trp Ser Phe Val Phe
 65 70 75
 15 Glu Asp Phe Val Ser Asp Glu Leu Arg Asn Lys Ala Thr Gln Pro Met
 80 85 90
 Lys Val Lys Phe Thr His Gly Gly Thr Gly Ser Ser Gln Thr Ala Pro
 95 100 105
 20 Thr Cys Gly Arg Glu Ser Ser Pro Arg Glu Thr Lys Leu Arg Met Ala
 110 115 120
 Ser Met Glu Ser Pro Gln
 125 130

<210> 385
 25 <211> 354
 <212> PRT
 <213> Homo sapiens

<400> 385
 30 Met Ser Ala Gly Gly Gly Arg Ala Phe Ala Trp Gln Val Phe Pro Pro
 1 5 10 15
 Met Pro Thr Cys Arg Val Tyr Gly Thr Val Ala His Gln Asp Gly His
 20 25 30
 35 Leu Leu Val Leu Gly Gly Cys Gly Arg Ala Gly Leu Pro Leu Asp Thr
 35 40 45
 Ala Glu Thr Leu Asp Met Ala Ser His Thr Trp Leu Ala Leu Ala Pro
 50 55 60
 Leu Pro Thr Ala Arg Ala Gly Ala Ala Val Val Leu Gly Lys Gln
 65 70 75 80
 40 Val Leu Val Val Cys Gly Val Asp Glu Val Gln Ser Pro Val Ala Ala
 85 90 95
 Val Glu Ala Phe Leu Met Asp Glu Gly Arg Trp Glu Arg Arg Ala Thr
 100 105 110
 45 Leu Pro Gln Ala Ala Met Gly Val Ala Thr Val Glu Arg Asp Gly Met
 115 120 125
 Val Tyr Ala Leu Gly Gly Met Gly Pro Asp Thr Ala Pro Gln Ala Gln
 130 135 140
 Val Arg Val Tyr Asp Pro Arg Arg Asp Cys Trp Leu Ser Leu Pro Ser
 145 150 155 160
 50 Met Pro Thr Pro Cys Tyr Gly Ala Ser Thr Phe Leu His Gly Asn Lys
 165 170 175
 Ile Tyr Val Leu Gly Gly Arg Gln Gly Lys Leu Pro Val Thr Ala Phe
 180 185 190
 Glu Ala Phe Asp Leu Glu Ala Arg Thr Trp Thr Arg His Pro Ser Leu
 195 200 205
 55 Pro Ser Arg Arg Ala Phe Ala Gly Cys Ala Met Ala Glu Gly Ser Val
 210 215 220
 Phe Ser Leu Gly Gly Leu Gln Gln Pro Gly Pro His Asn Phe Tyr Ser
 225 230 235 240
 60 Arg Pro His Phe Val Asn Thr Val Glu Met Phe Asp Leu Glu His Gly
 245 250 255
 Ser Trp Thr Lys Leu Pro Arg Ser Leu Arg Met Arg Asp Lys Arg Ala
 260 265 270

Asp Phe Val Val Gly Ser Leu Gly Gly His Ile Val Ala Ile Gly Gly
 275 280 285
 Leu Gly Asn Gln Pro Cys Pro Leu Gly Ser Val Glu Ser Phe Ser Leu
 290 295 300
 5 Ala Arg Arg Arg Trp Glu Ala Leu Pro Ala Met Pro Thr Ala Arg Cys
 305 310 315 320
 Ser Cys Ser Ser Leu Gln Ala Gly Pro Arg Leu Phe Val Ile Gly Gly
 325 330 335
 Val Ala Gln Gly Pro Ser Gln Ala Val Glu Ala Leu Cys Leu Arg Asp
 10 340 345 350
 Gly Val

<210> 386

<211> 207

15 <212> PRT

<213> Homo sapiens

<400> 386

Met Ala Leu Leu Phe Ala Arg Ser Leu Arg Leu Cys Arg Trp Gly Ala
 1 5 10 15
 Lys Arg Leu Gly Val Ala Ser Thr Glu Ala Gln Arg Gly Val Ser Phe
 20 25 30
 Lys Leu Glu Glu Lys Thr Ala His Ser Ser Leu Ala Leu Phe Arg Asp
 35 40 45
 25 Asp Thr Gly Val Lys Tyr Gly Leu Val Gly Leu Glu Pro Thr Lys Val
 50 55 60
 Ala Leu Asn Val Glu Arg Phe Arg Glu Trp Ala Val Val Leu Ala Asp
 65 70 75 80
 Thr Ala Val Thr Ser Gly Arg His Tyr Trp Glu Val Thr Val Lys Arg
 85 90 95
 30 Ser Gln Gln Phe Arg Ile Gly Val Ala Asp Val Asp Met Ser Arg Asp
 100 105 110
 Ser Cys Ile Gly Val Asp Asp Arg Ser Trp Val Phe Thr Tyr Ala Gln
 115 120 125
 35 Arg Lys Trp Tyr Thr Met Leu Ala Asn Glu Lys Ala Pro Val Glu Gly
 130 135 140
 Ile Gly Gln Pro Glu Lys Val Gly Leu Leu Leu Glu Tyr Glu Ala Gln
 145 150 155 160
 Lys Leu Ser Leu Val Asp Val Ser Gln Val Ser Val Val His Thr Leu
 165 170 175
 40 Gln Thr Asp Phe Arg Gly Pro Val Val Pro Ala Phe Ala Leu Trp Asp
 180 185 190
 Gly Glu Leu Leu Thr His Ser Gly Leu Glu Val Pro Glu Gly Leu
 195 200 205

45

<210> 387

<211> 210

<212> PRT

<213> Homo sapiens

50

<400> 387

Met Ala Ala Ser Val Glu Gln Arg Glu Gly Thr Ile Gln Val Gln Gly
 1 5 10 15
 Gln Ala Leu Phe Phe Arg Glu Ala Leu Pro Gly Ser Gly Gln Ala Arg
 20 25 30
 Phe Ser Val Leu Leu Leu His Gly Ile Arg Phe Ser Ser Glu Thr Trp
 35 40 45
 Gln Asn Leu Gly Thr Leu His Arg Leu Ala Gln Ala Gly Tyr Arg Ala
 50 55 60
 60 Val Ala Ile Asp Leu Pro Gly Leu Gly His Ser Lys Glu Ala Ala Ala
 65 70 75 80
 Pro Ala Pro Ile Gly Glu Leu Ala Pro Gly Ser Phe Leu Ala Ala Val
 85 90 95

Val Asp Ala Leu Glu Leu Gly Pro Pro Val Val Ile Ser Pro Ser Leu
 100 105 110
 Ser Gly Met Tyr Ser Leu Pro Phe Leu Thr Ala Pro Gly Ser Gln Leu
 115 120 125
 5 Pro Gly Phe Val Pro Val Ala Pro Ile Cys Thr Asp Lys Ile Asn Ala
 130 135 140
 Ala Asn Tyr Ala Ser Val Lys Thr Pro Ala Leu Ile Val Tyr Gly Asp
 145 150 155 160
 10 Gln Asp Pro Met Gly Gln Thr Ser Phe Glu His Leu Lys Gln Leu Pro
 165 170 175
 Asn His Arg Val Leu Ile Met Lys Gly Ala Gly His Pro Cys Tyr Leu
 180 185 190
 Asp Lys Pro Glu Glu Trp His Thr Gly Leu Leu Asp Phe Leu Gln Gly
 195 200 205
 15 Leu Gln
 210

 <210> 388
 <211> 375
 20 <212> PRT
 <213> Homo sapiens

 <400> 388
 25 Met Ala Val Thr Glu Ala Ser Leu Leu Arg Gln Cys Pro Leu Leu Leu
 1 5 10 15
 Pro Gln Asn Arg Ser Lys Thr Val Tyr Glu Gly Phe Ile Ser Ala Gln
 20 25 30
 Gly Arg Asp Phe His Leu Arg Ile Val Leu Pro Glu Asp Leu Gln Leu
 35 40 45
 30 Lys Asn Ala Arg Leu Leu Cys Ile Trp Gln Leu Arg Thr Ile Leu Ser
 50 55 60
 Gly Tyr His Arg Ile Val Gln Gln Arg Met Gln His Ser Pro Asp Leu
 65 70 75 80
 35 Met Ser Phe Met Met Glu Leu Lys Met Leu Leu Glu Val Ala Leu Lys
 85 90 95
 Asn Arg Gln Glu Leu Tyr Ala Leu Pro Pro Pro Gln Phe Tyr Ser
 100 105 110
 Ser Leu Ile Glu Glu Ile Gly Thr Leu Gly Trp Asp Lys Leu Val Tyr
 115 120 125
 40 Ala Asp Thr Cys Phe Ser Thr Ile Lys Leu Lys Ala Glu Asp Ala Ser
 130 135 140
 Gly Arg Glu His Leu Ile Thr Leu Lys Leu Lys Ala Lys Tyr Pro Ala
 145 150 155 160
 45 Glu Ser Pro Asp Tyr Phe Val Asp Phe Pro Val Pro Phe Cys Ala Ser
 165 170 175
 Trp Thr Pro Gln Ser Ser Leu Ile Ser Ile Tyr Ser Gln Phe Leu Ala
 180 185 190
 Ala Ile Glu Ser Leu Lys Ala Phe Trp Asp Val Met Asp Glu Ile Asp
 195 200 205
 50 Glu Lys Thr Trp Val Leu Glu Pro Glu Lys Pro Pro Arg Ser Ala Thr
 210 215 220
 Ala Arg Arg Ile Ala Leu Gly Asn Asn Val Ser Ile Asn Ile Glu Val
 225 230 235 240
 Asp Pro Arg His Pro Thr Met Leu Pro Glu Cys Phe Phe Leu Gly Ala
 245 250 255
 55 Asp His Val Val Lys Pro Leu Gly Ile Lys Leu Ser Arg Asn Ile His
 260 265 270
 Leu Trp Asp Pro Glu Asn Ser Val Leu Gln Asn Leu Lys Asp Val Leu
 275 280 285
 60 Glu Ile Asp Phe Pro Ala Arg Ala Ile Leu Glu Lys Ser Asp Phe Thr
 290 295 300
 Met Asp Cys Gly Ile Cys Tyr Ala Tyr Gln Leu Asp Gly Thr Ile Pro
 305 310 315 320

Asp Gln Val Cys Asp Asn Ser Gln Cys Gly Gln Pro Phe His Gln Ile
 325 330 335
 Cys Leu Tyr Glu Trp Leu Arg Gly Leu Leu Thr Ser Arg Gln Ser Phe
 340 345 350
 5 Asn Ile Ile Phe Gly Glu Cys Pro Tyr Cys Ser Lys Pro Ile Thr Leu
 355 360 365
 Lys Met Ser Gly Arg Lys His
 370 375
 10 <210> 389
 <211> 509
 <212> PRT
 <213> Homo sapiens
 15 <400> 389
 Met Ala Ala Ile Gly Val His Leu Gly Cys Thr Ser Ala Cys Val Ala
 1 5 10 15
 Val Tyr Lys Asp Gly Arg Ala Gly Val Val Ala Asn Asp Ala Gly Asp
 20 20 25 30
 Arg Val Thr Pro Ala Val Val Ala Tyr Ser Glu Asn Glu Ile Val
 35 40 45
 Gly Leu Ala Ala Lys Gln Ser Arg Ile Arg Asn Ile Ser Asn Thr Val
 50 55 60
 Met Lys Val Lys Gln Ile Leu Gly Arg Ser Ser Ser Asp Pro Gln Ala
 25 65 70 75 80
 Gln Lys Tyr Ile Ala Glu Ser Lys Cys Leu Val Ile Glu Lys Asn Gly
 85 90 95
 Lys Leu Arg Tyr Glu Ile Asp Thr Gly Glu Glu Thr Lys Phe Val Asn
 100 105 110
 30 Pro Glu Asp Val Ala Arg Leu Ile Phe Ser Lys Met Lys Glu Thr Ala
 115 120 125
 His Ser Val Leu Gly Ser Asp Ala Asn Asp Val Val Ile Thr Val Pro
 130 135 140
 Phe Asp Phe Gly Glu Lys Gln Lys Asn Ala Leu Gly Glu Ala Ala Arg
 35 145 150 155 160
 Ala Ala Gly Phe Asn Val Leu Arg Leu Ile His Glu Pro Ser Ala Ala
 165 170 175
 Leu Leu Ala Tyr Gly Ile Gly Gln Asp Ser Pro Thr Gly Lys Ser Asn
 180 185 190
 40 Ile Leu Val Phe Lys Leu Gly Gly Thr Ser Leu Ser Leu Ser Val Met
 195 200 205
 Glu Val Asn Ser Gly Ile Tyr Arg Val Leu Ser Thr Asn Thr Asp Asp
 210 215 220
 Asn Ile Gly Gly Ala His Phe Thr Glu Thr Leu Ala Gln Tyr Leu Ala
 45 225 230 235 240
 Ser Glu Phe Gln Arg Ser Phe Lys His Asp Val Arg Gly Asn Ala Arg
 245 250 255
 Ala Met Met Lys Leu Thr Asn Ser Ala Glu Val Ala Lys His Ser Leu
 260 265 270
 50 Ser Thr Leu Gly Ser Ala Asn Cys Phe Leu Asp Ser Leu Tyr Glu Gly
 275 280 285
 Gln Asp Phe Asp Cys Asn Val Ser Arg Ala Arg Phe Glu Leu Leu Cys
 290 295 300
 Ser Pro Leu Phe Asn Lys Cys Ile Glu Ala Ile Arg Gly Leu Leu Asp
 55 305 310 315 320
 Gln Asn Gly Phe Thr Thr Asp Asp Ile Asn Lys Val Val Leu Cys Gly
 325 330 335
 Gly Ser Ser Arg Ile Pro Lys Leu Gln Gln Leu Ile Lys Asp Leu Phe
 340 345 350
 60 Pro Ala Val Glu Leu Leu Asn Ser Ile Pro Pro Asp Glu Val Ile Pro
 355 360 365
 Ile Gly Ala Ala Ile Glu Ala Gly Ile Leu Ile Gly Lys Glu Asn Leu
 370 375 380

Leu Val Glu Asp Ser Leu Met Ile Glu Cys Ser Ala Arg Asp Ile Leu
 385 390 395 400
 Val Lys Gly Val Asp Glu Ser Gly Ala Ser Arg Phe Thr Val Leu Phe
 405 410 415
 5 Pro Ser Gly Thr Pro Leu Pro Ala Arg Arg Gln His Thr Leu Gln Ala
 420 425 430
 Pro Gly Ser Ile Ser Ser Val Cys Leu Glu Leu Tyr Glu Ser Asp Gly
 435 440 445
 Lys Asn Ser Ala Lys Glu Glu Thr Lys Phe Ala Gln Val Val Leu Gln
 10 450 455 460
 Asp Leu Asp Lys Lys Glu Asn Gly Leu Arg Asp Ile Leu Ala Val Leu
 465 470 475 480
 Thr Met Lys Arg Asp Gly Ser Leu His Val Thr Cys Thr Asp Gln Glu
 485 490 495
 15 Thr Gly Lys Cys Glu Ala Ile Ser Ile Glu Ile Ala Ser
 500 505

<210> 390

<211> 78

20 <212> PRT

<213> Homo sapiens

<400> 390

Met Tyr Asn Thr Gly Arg His Val Ser Leu Arg Leu Asp Lys Glu His
 25 1 5 10 15
 Leu Val Asn Ile Ser Gly Gly Pro Met Thr Tyr Ser His Arg Leu Glu
 20 25 30
 Glu Ile Arg Leu His Phe Gly Ser Glu Asp Ser Gln Gly Ser Glu His
 35 40 45
 30 Leu Leu Asn Gly Gln Ala Phe Ser Gly Glu Leu Gln Glu Arg Asp Leu
 50 55 60
 Phe Ile Leu Leu Thr Ser Val Ser Gly His Leu Pro Asp Thr
 65 70 75

35 <210> 391

<211> 162

<212> PRT

<213> Homo sapiens

40 <400> 391

Met Ala Thr His Ala Leu Glu Ile Ala Gly Leu Phe Leu Gly Gly Val
 1 5 10 15
 Gly Met Val Gly Thr Val Ala Val Thr Val Met Pro Gln Trp Ile Val
 20 25 30
 45 Ser Ala Phe Ile Glu Asn Asn Ile Val Val Phe Glu Asn Phe Trp Glu
 35 40 45
 Gly Leu Trp Met Asn Cys Val Arg Gln Ala Asn Ile Arg Met Gln Cys
 50 55 60
 Lys Ile Tyr Asp Ser Leu Leu Ala Leu Ser Pro Asp Leu Gln Ala Ala
 50 65 70 75 80
 Arg Gly Leu Met Cys Ala Ala Ser Val Met Ser Phe Leu Ala Phe Met
 85 90 95
 Met Ala Ile Leu Gly Met Lys Cys Thr Arg Cys Thr Gly Asp Asn Glu
 100 105 110
 55 Lys Val Lys Ala His Ile Leu Leu Thr Ala Gly Ile Ile Phe Ile Ile
 115 120 125
 Thr Gly Met Val Val Leu Ile Pro Val Ser Trp Val Ala Asn Ala Ile
 130 135 140
 Ile Arg Asp Phe Tyr Asn Pro Ile Val Asn Val Ala Gln Lys Arg Glu
 60 145 150 155 160
 Leu Gly

<210> 392

<211> 146
 <212> PRT
 <213> Homo sapiens

5 <400> 392
 Met Asn Ser Leu Leu His Phe Gly Ile Leu Leu Glu Leu Ser Leu Leu
 1 5 10 15
 Lys Gln Phe Lys Ser Val Tyr Val Pro Gly Asn His Thr His Gln Ala
 20 25 30
 10 Ser Tyr Lys Pro Leu Leu Lys Gln Val Val Glu Glu Ile Phe His Pro
 35 40 45
 Glu Arg Pro Asp Ser Val Asp Ile Glu His Met Ser Ser Gly Leu Thr
 50 55 60
 Asp Leu Leu Lys Thr Gly Phe Ser Met Phe Met Lys Val Ser Arg Pro
 15 65 70 75 80
 His Pro Ser Asp Tyr Pro Leu Leu Ile Leu Phe Val Val Gly Gly Val
 85 90 95
 Thr Val Ser Glu Val Lys Met Val Lys Asp Leu Val Ala Ser Leu Lys
 100 105 110
 20 Pro Gly Thr Gln Val Ile Val Leu Ser Thr Arg Leu Leu Lys Pro Leu
 115 120 125
 Asn Ile Pro Glu Leu Leu Phe Ala Thr Asp Arg Leu His Pro Asp Leu
 130 135 140
 Gly Phe
 25 145

<210> 393
 <211> 225
 <212> PRT

30 <213> Homo sapiens

<400> 393
 Met Ala Thr His Ala Leu Glu Ile Ala Gly Leu Phe Leu Gly Gly Val
 1 5 10 15
 35 Gly Met Val Gly Thr Val Ala Val Thr Val Met Pro Gln Trp Arg Val
 20 25 30
 Ser Ala Phe Ile Glu Asn Asn Ile Val Val Phe Glu Asn Phe Trp Glu
 35 40 45
 Gly Leu Trp Met Asn Cys Val Arg Gln Ala Asn Ile Arg Met Gln Cys
 50 55 60
 40 Lys Ile Tyr Asp Ser Leu Leu Ala Leu Ser Pro Asp Leu Gln Ala Ala
 65 70 75 80
 Arg Gly Leu Met Cys Ala Ala Ser Val Met Ser Phe Leu Ala Phe Met
 85 90 95
 45 Met Ala Ile Leu Gly Met Lys Cys Thr Arg Cys Thr Gly Asp Asn Glu
 100 105 110
 Lys Val Lys Ala His Ile Leu Leu Thr Ala Gly Ile Ile Phe Ile Ile
 115 120 125
 Ala Gly Met Val Val Leu Ile Pro Val Ser Trp Val Ala Asn Ala Ile
 130 135 140
 50 Ile Arg Asp Phe Tyr Asn Pro Ile Val Asn Val Ala Gln Lys Arg Glu
 145 150 155 160
 Leu Gly Glu Ala Leu Tyr Leu Gly Trp Thr Thr Ala Leu Val Leu Ile
 165 170 175
 55 Val Gly Gly Ala Leu Phe Cys Cys Val Phe Cys Cys Asn Glu Lys Ser
 180 185 190
 Ser Ser Tyr Arg Tyr Ser Ile Pro Ser His Arg Thr Thr Gln Lys Ser
 195 200 205
 Tyr His Thr Gly Lys Lys Ser Pro Ser Val Tyr Ser Arg Ser Gln Tyr
 210 215 220
 60 Val
 225

<210> 394
 <211> 114
 <212> PRT
 <213> Homo sapiens

5

<400> 394

Met Arg Leu Gln Asp Arg Ile Ala Thr Phe Phe Phe Pro Lys Gly Met
 1 5 10 15
 Met Leu Thr Thr Ala Ala Leu Met Leu Phe Phe Leu His Leu Gly Ile
 10 20 25 30
 Phe Ile Arg Asp Val His Asn Phe Cys Ile Thr Tyr His Tyr Asp His
 35 40 45
 Met Ser Phe His Tyr Thr Val Val Leu Met Phe Ser Gln Val Ile Ser
 50 55 60
 Ile Cys Trp Ala Ala Met Gly Ser Leu Tyr Ala Glu Met Thr Glu Asn
 15 65 70 75 80
 Asn Ala Gln Arg Ser His Val Leu Gln Pro Pro Val Leu Gly Val Ser
 85 90 95
 Gly His Arg Val Pro Gly Gly Ala Pro Leu Arg Pro Gly Glu Ser Glu
 20 100 105 110
 Gln Gly

<210> 395

<211> 367

25 <212> PRT

<213> Homo sapiens

<400> 395

Met Ala Thr Pro Asn Asn Leu Thr Pro Thr Asn Cys Ser Trp Trp Pro
 1 5 10 15
 Ile Ser Ala Leu Glu Ser Asp Ala Ala Lys Pro Ala Glu Ala Pro Asp
 20 25 30
 Ala Pro Glu Ala Ala Ser Pro Ala His Trp Pro Arg Glu Ser Leu Val
 35 35 40 45
 Leu Tyr His Trp Thr Gln Ser Phe Ser Ser Gln Lys Val Arg Leu Val
 50 55 60
 Ile Ala Glu Lys Gly Leu Val Cys Glu Glu Arg Asp Val Ser Leu Pro
 65 70 75 80
 Gln Ser Glu His Lys Glu Pro Trp Phe Met Arg Leu Asn Leu Gly Glu
 40 85 90 95
 Glu Val Pro Val Ile Ile His Arg Asp Asn Ile Ile Ser Asp Tyr Asp
 100 105 110
 Gln Ile Ile Asp Tyr Val Glu Arg Thr Phe Thr Gly Glu His Val Val
 115 120 125
 Ala Leu Met Pro Glu Val Gly Ser Leu Gln His Ala Arg Val Leu Gln
 130 135 140
 Tyr Arg Glu Leu Leu Asp Ala Leu Pro Met Asp Ala Tyr Thr His Gly
 145 150 155 160
 Cys Ile Leu His Pro Glu Leu Thr Thr Asp Ser Met Ile Pro Lys Tyr
 165 170 175
 Ala Thr Ala Glu Ile Arg Arg His Leu Ala Asn Ala Thr Thr Asp Leu
 180 185 190
 Met Lys Leu Asp His Glu Glu Glu Pro Gln Leu Ser Glu Pro Tyr Leu
 195 200 205
 Ser Lys Gln Lys Lys Leu Met Val Lys Ile Leu Glu His Asp Asp Val
 210 215 220
 Ser Tyr Leu Lys Lys Ile Leu Gly Glu Leu Ala Met Val Leu Asp Gln
 225 230 235 240
 Ile Glu Ala Glu Leu Glu Lys Arg Lys Leu Glu Asn Glu Gly Gln Lys
 245 250 255
 Cys Glu Leu Trp Leu Cys Gly Cys Ala Phe Thr Leu Ala Asp Val Leu
 260 265 270
 Leu Gly Ala Thr Leu His Arg Leu Lys Phe Leu Gly Leu Ser Lys Lys

275 280 285
 Tyr Trp Glu Asp Gly Ser Arg Pro Asn Leu Gln Ser Phe Phe Glu Arg
 290 295 300
 Val Gln Arg Arg Phe Ala Phe Arg Lys Val Leu Gly Asp Ile His Thr
 5 305 310 315 320
 Thr Leu Leu Ser Ala Val Ile Pro Asn Ala Phe Arg Leu Val Lys Arg
 325 330 335
 Lys Pro Pro Ser Phe Phe Gly Ala Ser Phe Leu Met Gly Ser Leu Gly
 340 345 350
 10 Gly Met Gly Tyr Phe Ala Tyr Trp Tyr Leu Lys Lys Lys Tyr Ile
 355 360 365

<210> 396
 <211> 279
 15 <212> PRT
 <213> Homo sapiens

<400> 396
 Met Pro Val Cys Ala Pro Val Leu Trp Arg Ala Arg Arg Leu Cys Gly
 1 5 10 15
 Met Pro Val Cys Ala Pro Val Pro Trp Arg Ala Arg Arg Leu Cys Thr
 20 20 25 30
 Arg Ala Val Val Cys Pro Ser Ser Val Pro Phe Ile Ala Gly Gln Gly
 35 40 45
 25 Cys Thr His Met Cys Lys Pro Ala Thr Asp Pro Arg Phe Thr Arg Ser
 50 55 60
 Pro Leu Ala Gly Gly Val Ile Leu Gly Val Ala Leu Trp Leu Arg His
 65 70 75 80
 Asp Pro Gln Thr Thr Asn Leu Leu Tyr Leu Glu Leu Gly Asp Lys Pro
 30 85 90 95
 Ala Pro Asn Thr Phe Tyr Val Gly Ile Tyr Ile Leu Ile Ala Val Gly
 100 105 110
 Ala Val Met Met Phe Val Gly Phe Leu Gly Cys Tyr Gly Ala Ile Gln
 115 120 125
 35 Glu Ser Gln Cys Leu Leu Gly Thr Phe Phe Thr Cys Leu Val Ile Leu
 130 135 140
 Phe Ala Cys Glu Val Ala Ala Gly Ile Trp Gly Phe Val Asn Lys Asp
 145 150 155 160
 Gln Ile Ala Lys Asp Val Lys Gln Phe Tyr Asp Gln Ala Leu Gln Gln
 40 165 170 175
 Ala Val Val Asp Asp Asp Ala Asn Asn Ala Lys Ala Val Val Lys Thr
 180 185 190
 Phe His Glu Thr Leu Asp Cys Cys Gly Ser Ser Thr Leu Thr Ala Leu
 195 200 205
 45 Thr Thr Ser Val Leu Lys Asn Asn Leu Cys Pro Ser Gly Ser Asn Ile
 210 215 220
 Ile Ser Asn Leu Phe Lys Glu Asp Cys His Gln Lys Ile Asp Asp Leu
 225 230 235 240
 Phe Ser Gly Lys Leu Tyr Leu Ile Gly Ile Ala Ala Ile Val Val Ala
 50 245 250 255
 Val Ile Met Ile Phe Glu Met Ile Leu Ser Met Val Leu Cys Cys Gly
 260 265 270
 Ile Arg Asn Ser Ser Val Tyr
 275

55
 <210> 397
 <211> 173
 <212> PRT
 <213> Homo sapiens

60
 <400> 397
 Met Cys Leu Leu Leu Gly Ala Thr Gly Val Gly Lys Thr Leu Leu Val
 1 5 10 15

Lys Arg Leu Gln Glu Val Ser Ser Arg Asp Gly Lys Gly Asp Leu Gly
 20 25 30
 Glu Pro Pro Pro Thr Arg Pro Thr Val Gly Thr Asn Leu Thr Asp Ile
 35 40 45
 5 Val Ala Gln Arg Lys Ile Thr Ile Arg Glu Leu Gly Gly Cys Met Gly
 50 55 60
 Pro Ile Trp Ser Ser Tyr Tyr Gly Asn Cys Arg Ser Leu Leu Phe Val
 65 70 75 80
 10 Met Asp Ala Ser Asp Pro Thr Gln Leu Ser Ala Ser Cys Val Gln Leu
 85 90 95
 Leu Gly Leu Leu Ser Ala Glu Gln Leu Ala Glu Ala Ser Val Leu Ile
 100 105 110
 Leu Phe Asn Lys Ile Asp Leu Pro Cys Tyr Met Ser Thr Glu Glu Met
 115 120 125
 15 Lys Ser Leu Ile Arg Leu Pro Asp Ile Ile Ala Cys Ala Lys Gln Asn
 130 135 140
 Ile Thr Thr Ala Glu Ile Ser Ala Arg Glu Gly Thr Gly Leu Ala Gly
 145 150 155 160
 20 Val Leu Ala Trp Leu Gln Ala Thr His Arg Ala Asn Asp
 165 170

<210> 398

<211> 205

<212> PRT

25 <213> Homo sapiens

<400> 398

Met Ala Ala Ala Arg Pro Ser Leu Gly Arg Val Leu Pro Gly Ser Ser
 1 5 10 15
 30 Val Leu Phe Leu Cys Asp Met Gln Glu Lys Phe Arg His Asn Ile Ala
 20 25 30
 Tyr Phe Pro Gln Ile Val Ser Val Ala Ala Arg Met Leu Lys Val Ala
 35 40 45
 35 Arg Leu Leu Glu Val Pro Val Met Leu Thr Glu Gln Tyr Pro Gln Gly
 50 55 60
 Leu Gly Pro Thr Val Pro Glu Leu Gly Thr Glu Gly Leu Arg Pro Leu
 65 70 75 80
 Ala Lys Thr Cys Phe Ser Met Val Pro Ala Leu Gln Gln Glu Leu Asp
 85 90 95
 40 Ser Arg Pro Gln Leu Arg Ser Val Leu Leu Cys Gly Ile Glu Ala Gln
 100 105 110
 Ala Cys Ile Leu Asn Thr Thr Leu Asp Leu Leu Asp Arg Gly Leu Gln
 115 120 125
 45 Val His Val Val Val Asp Ala Cys Ser Ser Arg Ser Gln Val Asp Arg
 130 135 140
 Leu Val Ala Leu Ala Arg Met Arg Gln Ser Gly Ala Phe Leu Ser Thr
 145 150 155 160
 Ser Glu Gly Leu Ile Leu Gln Leu Val Gly Asp Ala Val His Pro Gln
 165 170 175
 50 Phe Lys Glu Ile Gln Lys Leu Ile Lys Glu Pro Ala Pro Asp Ser Gly
 180 185 190
 Leu Leu Gly Leu Phe Gln Gly Gln Asn Ser Leu Leu His
 195 200 205

55 <210> 399

<211> 180

<212> PRT

<213> Homo sapiens

60 <400> 399

Met Trp Leu Tyr Arg Asn Pro Tyr Val Glu Ala Glu Tyr Phe Pro Thr
 1 5 10 15
 Lys Pro Met Phe Val Ile Ala Phe Leu Ser Pro Leu Ser Leu Ile Phe

20 25 30
 Leu Ala Lys Phe Leu Lys Lys Ala Asp Thr Arg Asp Ser Arg Gln Ala
 35 40 45
 Cys Leu Ala Ala Ser Leu Ala Leu Ala Leu Asn Gly Val Phe Thr Asn
 5 50 55 60
 Thr Ile Lys Leu Ile Val Gly Arg Pro Arg Pro Asp Phe Phe Tyr Arg
 65 70 75 80
 Cys Phe Pro Asp Gly Leu Ala His Ser Asp Leu Met Cys Thr Gly Asp
 85 90 95
 10 Lys Asp Val Val Asn Glu Gly Arg Lys Ser Phe Pro Ser Gly His Ser
 100 105 110
 Ser Phe Ala Phe Ala Gly Leu Ala Phe Ala Ser Phe Tyr Leu Ala Gly
 115 120 125
 Lys Leu His Cys Phe Thr Pro Gln Gly Arg Gly Lys Ser Trp Arg Phe
 15 130 135 140
 Cys Ala Phe Leu Ser Pro Leu Leu Phe Ala Ala Val Ile Ala Leu Ser
 145 150 155 160
 Arg Thr Cys Asp Tyr Lys His His Trp Gln Asp Leu Leu Lys Cys Thr
 165 170 175
 20 Asn Thr Ala Lys
 180

<210> 400

<211> 150

25 <212> PRT

<213> Homo sapiens

<400> 400

30 Met Cys Thr Ala Leu Leu Leu Leu Tyr Leu Arg Trp Cys Phe Asn Leu
 1 5 10 15
 Lys Leu Val Asn Val Lys Tyr Glu Pro Lys Asp Ser Leu Gly Pro Glu
 20 25 30
 Met Thr Phe Val Ala Asp Ala Ala Arg Gly Pro Leu Leu Ser Ser Leu
 35 35 40 45
 Asp Ser Pro Ala Asn Leu Met Ser Thr Ala Ser Val Cys Ile Ser Leu
 50 55 60
 Pro Glu Gly Cys Ser Gly Gly Arg Ser Pro Cys Tyr Ser Gln Lys Trp
 65 70 75 80
 Pro Pro Glu Val Pro Glu Lys Leu Thr Ser Leu Gly Gln Gln Ser Ser
 40 85 90 95
 Thr Ser Ser Leu Thr Asp Thr Asp Val Gln Val Ser Pro Met Leu Val
 100 105 110
 Ala Gly Val Asn His Ser Ser Ser Leu Leu Asp Asn Ile Pro Phe Thr
 115 120 125
 45 Gly Cys Leu Pro Phe His Leu Ser Ser Ser Leu Pro Tyr Leu Cys Leu
 130 135 140
 Leu Gly Ser Pro Phe Lys
 145 150

50 <210> 401

<211> 170

<212> PRT

<213> Homo sapiens

55 <400> 401

Met Glu Asp Pro Asn Pro Glu Glu Asn Met Lys Gln Gln Asp Ser Pro
 1 5 10 15
 Lys Glu Arg Ser Pro Gln Ser Pro Gly Gly Asn Ile Cys His Leu Gly
 20 25 30
 60 Ala Pro Lys Cys Thr Arg Cys Leu Ile Thr Phe Ala Asp Ser Lys Phe
 35 40 45
 Gln Glu Arg His Met Lys Arg Glu His Pro Ala Asp Phe Val Ala Gln
 50 55 60

Lys Leu Gln Gly Val Leu Phe Ile Cys Phe Thr Cys Ala Arg Ser Phe
 65 70 75 80
 Pro Ser Ser Lys Ala Leu Ile Thr His Gln Arg Ser His Gly Pro Ala
 85 90 95
 5 Ala Lys Pro Thr Leu Pro Val Ala Thr Thr Thr Ala Gln Pro Thr Phe
 100 105 110
 Pro Cys Pro Asp Cys Gly Lys Thr Phe Gly Gln Ala Val Ser Leu Arg
 115 120 125
 Arg His Arg Gln Met His Glu Val Arg Ala Pro Pro Gly Thr Phe Ala
 10 130 135 140
 Cys Thr Glu Cys Gly Gln Asp Phe Ala Gln Glu Ala Gly Leu His Gln
 145 150 155 160
 His Tyr Ile Arg His Ala Arg Gly Glu Leu
 165 170
 15
 <210> 402
 <211> 169
 <212> PRT
 <213> Homo sapiens
 20
 <400> 402
 Met Glu Asp Pro Asn Pro Glu Glu Asn Met Lys Gln Gln Asp Ser Pro
 1 5 10 15
 Lys Glu Arg Ser Pro Gln Pro Arg Arg Gln His Leu Pro Pro Gly Gly
 25 20 25 30
 Pro Glu Val His Pro Leu Pro His His Leu Arg Arg Phe Gln Val Pro
 35 40 45
 Gly Ala Ser His Glu Ala Gly Ala Pro Ser Gly Leu Arg Gly Pro Glu
 50 55 60
 30 Ala Ala Gly Gly Pro Leu His Leu Leu His Leu Arg Pro Leu Leu Pro
 65 70 75 80
 Leu Leu Gln Ser Pro Asn His Pro Pro Ala Gln His Gly Pro Ala Ala
 85 90 95
 Lys Pro Thr Leu Pro Val Ala Thr Thr Thr Ala Gln Pro Thr Phe Pro
 35 100 105 110
 Cys Pro Asp Cys Gly Lys Thr Phe Gly Gln Ala Val Ser Leu Arg Arg
 115 120 125
 His Arg Gln Met His Glu Val Arg Ala Pro Pro Gly Thr Phe Ala Cys
 130 135 140
 40 Thr Glu Cys Gly Gln Asp Phe Ala Gln Glu Ala Gly Leu His Gln His
 145 150 155 160
 Tyr Ile Arg His Ala Arg Gly Glu Leu
 165
 45
 <210> 403
 <211> 367
 <212> PRT
 <213> Homo sapiens
 50
 <400> 403
 Met Ala Thr Pro Asn Asn Leu Thr Pro Thr Asn Cys Ser Trp Trp Pro
 1 5 10 15
 Ile Ser Ala Leu Glu Ser Asp Ala Ala Lys Pro Ala Glu Ala Pro Asp
 20 25 30
 55 Ala Pro Glu Ala Ala Ser Pro Ala His Trp Pro Arg Glu Ser Leu Val
 35 40 45
 Leu Tyr His Trp Thr Gln Ser Phe Ser Ser Gln Lys Val Arg Leu Val
 50 55 60
 Ile Ala Glu Lys Gly Leu Val Cys Glu Glu Arg Asp Val Ser Leu Pro
 60 65 70 75 80
 Gln Ser Glu His Lys Glu Pro Trp Phe Met Arg Leu Asn Leu Gly Glu
 85 90 95
 Glu Val Pro Val Ile Ile His Arg Asp Asn Ile Ile Ser Asp Tyr Asp

100 105 110
 Gln Ile Ile Asp Tyr Val Glu Arg Thr Phe Thr Gly Glu His Val Val
 115 120 125
 Ala Leu Met Pro Glu Val Gly Ser Leu Gln His Ala Arg Val Leu Gln
 5 130 135 140
 Tyr Arg Glu Leu Leu Asp Ala Leu Pro Met Asp Ala Tyr Thr His Gly
 145 150 155 160
 Cys Ile Leu His Leu Glu Leu Thr Thr Asp Ser Met Ile Pro Lys Tyr
 165 170 175
 10 Ala Thr Ala Glu Ile Arg Arg His Leu Ala Asn Ala Thr Thr Asp Leu
 180 185 190
 Met Lys Leu Asp His Glu Glu Glu Pro Gln Leu Ser Glu Pro Tyr Leu
 195 200 205
 Ser Lys Gln Lys Lys Leu Met Ala Lys Ile Leu Glu His Asp Asp Val
 15 210 215 220
 Ser Tyr Leu Lys Lys Ile Leu Gly Glu Leu Ala Met Val Leu Asp Gln
 225 230 235 240
 Ile Glu Ala Glu Leu Glu Lys Arg Lys Leu Glu Asn Glu Gly Gln Lys
 245 250 255
 20 Cys Glu Leu Trp Leu Cys Gly Cys Ala Phe Thr Leu Ala Asp Val Leu
 260 265 270
 Leu Gly Ala Thr Leu His Arg Leu Lys Phe Leu Gly Leu Ser Lys Lys
 275 280 285
 Tyr Trp Glu Asp Gly Ser Arg Pro Asn Leu Gln Ser Phe Phe Glu Arg
 25 290 295 300
 Val Gln Arg Arg Phe Ala Phe Arg Lys Val Leu Gly Asp Ile His Thr
 305 310 315 320
 Thr Leu Leu Ser Ala Val Ile Pro Asn Ala Phe Arg Leu Val Lys Arg
 325 330 335
 30 Lys Pro Pro Ser Phe Phe Gly Ala Ser Phe Leu Met Gly Ser Leu Gly
 340 345 350
 Gly Met Gly Tyr Phe Ala Tyr Trp Tyr Leu Lys Lys Lys Tyr Ile
 355 360 365

 35 <210> 404
 <211> 20
 <212> PRT
 <213> Homo sapiens

 40 <400> 404
 Met Ala Ala Ala Arg Pro Ser Leu Gly Arg Val Leu Pro Gly Ser Ser
 1 5 10 15
 Pro Val Pro Val
 20
 45
 <210> 405
 <211> 225
 <212> PRT
 <213> Homo sapiens
 50
 <400> 405
 Met Ala Thr His Ala Leu Glu Ile Ala Gly Leu Phe Leu Gly Gly Val
 1 5 10 15
 Gly Met Val Gly Thr Val Ala Val Thr Val Met Pro Gln Trp Arg Val
 55 20 25 30
 Ser Ala Phe Ile Glu Asn Asn Ile Val Val Phe Glu Asn Phe Trp Glu
 35 40 45
 Gly Leu Trp Met Asn Cys Val Arg Gln Ala Asn Ile Arg Met Gln Cys
 50 55 60
 60 Lys Ile Tyr Asp Ser Leu Leu Ala Leu Ser Pro Asp Leu Gln Ala Ala
 65 70 75 80
 Arg Gly Leu Met Cys Ala Ala Ser Val Met Ser Phe Leu Ala Phe Met
 85 90 95

Met Ala Ile Leu Gly Met Lys Cys Thr Arg Cys Thr Gly Asp Asn Glu
 100 105 110
 Lys Val Lys Ala His Ile Leu Leu Thr Ala Gly Ile Ile Phe Ile Ile
 115 120 125
 5 Thr Gly Met Val Val Leu Ile Pro Val Ser Trp Val Ala Asn Ala Ile
 130 135 140
 Ile Arg Asp Phe Tyr Asn Ser Ile Val Asn Val Ala Gln Lys Arg Glu
 145 150 155 160
 10 Leu Gly Glu Ala Leu Tyr Leu Gly Trp Thr Thr Ala Leu Val Leu Ile
 165 170 175
 Val Gly Gly Ala Leu Phe Cys Cys Val Phe Cys Cys Asn Glu Lys Ser
 180 185 190
 Ser Ser Tyr Arg Tyr Ser Ile Pro Ser His Arg Thr Thr Gln Lys Ser
 195 200 205
 15 Tyr His Thr Gly Lys Lys Ser Pro Ser Val Tyr Ser Arg Ser Gln Tyr
 210 215 220
 Val
 225
 20 <210> 406
 <211> 378
 <212> PRT
 <213> Homo sapiens
 25 <400> 406
 Met Asp Pro Gly Asp Asp Trp Leu Val Glu Ser Leu Arg Leu Tyr Gln
 1 5 10 15
 Asp Phe Tyr Ala Phe Asp Leu Ser Gly Ala Thr Arg Val Leu Glu Trp
 20 25 30
 30 Ile Asp Asp Lys Gly Val Phe Val Ala Gly Tyr Glu Ser Leu Lys Lys
 35 40 45
 Asn Glu Ile Leu His Leu Lys Leu Pro Leu Arg Leu Ser Val Lys Glu
 50 55 60
 Asn Lys Gly Leu Phe Pro Glu Arg Asp Phe Lys Val Arg His Gly Gly
 35 65 70 75 80
 Phe Ser Asp Arg Ser Ile Phe Asp Leu Lys His Val Pro His Thr Arg
 85 90 95
 Leu Leu Val Thr Ser Gly Leu Pro Gly Cys Tyr Leu Gln Val Trp Gln
 100 105 110
 40 Val Ala Glu Asp Ser Asp Val Ile Lys Ala Val Ser Thr Ile Ala Val
 115 120 125
 His Glu Lys Glu Glu Ser Leu Trp Pro Arg Val Ala Val Phe Ser Thr
 130 135 140
 Leu Ala Pro Gly Val Leu His Gly Ala Arg Leu Arg Ser Leu Gln Val
 45 145 150 155 160
 Val Asp Leu Glu Ser Arg Lys Thr Thr Tyr Thr Ser Asp Val Ser Asp
 165 170 175
 Ser Glu Glu Leu Ser Ser Leu Gln Val Leu Asp Ala Asp Thr Phe Ala
 180 185 190
 50 Phe Cys Cys Ala Ser Gly Arg Leu Gly Leu Val Asp Thr Arg Gln Lys
 195 200 205
 Trp Ala Pro Leu Glu Asn Arg Ser Pro Gly Pro Gly Ser Gly Gly Glu
 210 215 220
 Arg Trp Cys Ala Glu Val Gly Ser Trp Gly Gln Gly Pro Gly Pro Ser
 55 225 230 235 240
 Ile Ala Ser Leu Ser Ser Asp Gly Arg Leu Cys Leu Leu Asp Pro Arg
 245 250 255
 Asp Leu Cys His Pro Val Ser Ser Val Gln Cys Pro Val Ser Val Pro
 260 265 270
 60 Ser Pro Asp Pro Glu Leu Leu Arg Val Thr Trp Ala Pro Gly Leu Lys
 275 280 285
 Asn Cys Leu Ala Ile Ser Gly Phe Asp Gly Thr Val Gln Val Tyr Asp
 290 295 300

Ala Thr Ser Trp Asp Gly Thr Arg Ser Gln Asp Gly Thr Arg Ser Gln
 305 310 315 320
 Val Glu Pro Leu Phe Thr His Arg Gly His Ile Phe Leu Asp Gly Asn
 325 330 335
 5 Gly Met Asp Pro Ala Pro Leu Val Thr Thr His Thr Trp His Pro Cys
 340 345 350
 Arg Pro Arg Thr Leu Leu Ser Ala Thr Asn Asp Ala Ser Leu His Val
 355 360 365
 Trp Asp Trp Val Asp Leu Cys Ala Pro Arg
 10 370 375

<210> 407
 <211> 43
 <212> PRT
 15 <213> Homo sapiens

<400> 407
 Met Ala Thr His Ala Leu Glu Ile Ala Gly Leu Phe Leu Gly Gly Val
 1 5 10 15
 20 Gly Met Val Gly Thr Val Ala Val Thr Val Met Pro Gln Trp Arg Val
 20 25 30
 Ser Ala Phe Ile Glu Asn Asn Ile Val Val Phe
 35 40

25 <210> 408
 <211> 345
 <212> PRT
 <213> Homo sapiens

30 <400> 408
 Met Ala Trp Arg Gly Trp Ala Gln Arg Gly Trp Gly Cys Gly Gln Ala
 1 5 10 15
 Trp Gly Ala Ser Val Gly Gly Arg Ser Cys Glu Glu Leu Thr Ala Val
 20 25 30
 35 Leu Thr Pro Pro Gln Leu Leu Gly Arg Arg Phe Asn Phe Phe Ile Gln
 35 40 45
 Gln Lys Cys Gly Phe Arg Lys Ala Pro Arg Lys Val Glu Pro Arg Arg
 50 55 60
 Ser Asp Pro Gly Thr Ser Gly Glu Ala Tyr Lys Arg Ser Ala Leu Ile
 40 65 70 75 80
 Pro Pro Val Glu Glu Thr Val Phe Tyr Pro Ser Pro Tyr Pro Ile Arg
 85 90 95
 Ser Leu Ile Lys Pro Leu Phe Phe Thr Val Gly Phe Thr Gly Cys Ala
 100 105 110
 45 Phe Gly Ser Ala Ala Ile Trp Gln Tyr Glu Ser Leu Lys Ser Arg Val
 115 120 125
 Gln Ser Tyr Phe Asp Gly Ile Lys Ala Asp Trp Leu Asp Ser Ile Arg
 130 135 140
 Pro Gln Lys Glu Gly Asp Phe Arg Lys Glu Ile Asn Lys Trp Trp Asn
 50 145 150 155 160
 Asn Leu Ser Asp Gly Gln Arg Thr Val Thr Gly Ile Ile Ala Ala Asn
 165 170 175
 Val Leu Val Phe Cys Leu Trp Arg Val Pro Ser Leu Gln Arg Thr Met
 180 185 190
 55 Ile Arg Tyr Phe Thr Ser Asn Pro Ala Ser Lys Val Leu Cys Ser Pro
 195 200 205
 Met Leu Leu Ser Thr Phe Ser His Phe Ser Leu Phe His Met Ala Ala
 210 215 220
 Asn Met Tyr Val Leu Trp Ser Phe Ser Ser Ser Ile Val Asn Ile Leu
 60 225 230 235 240
 Gly Gln Glu Gln Phe Met Ala Val Tyr Leu Ser Ala Gly Val Ile Ser
 245 250 255
 Asn Phe Val Ser Tyr Val Gly Lys Val Ala Thr Gly Arg Tyr Gly Pro

260 265 270
 Ser Leu Gly Ala Ala Leu Lys Ala Ile Ile Ala Met Asp Thr Ala Gly
 275 280 285
 Met Ile Leu Gly Trp Lys Phe Phe Asp His Ala Ala His Leu Gly Gly
 290 295 300
 5 Ala Leu Phe Gly Ile Trp Tyr Val Thr Tyr Gly His Glu Leu Ile Trp
 305 310 315 320
 Lys Asn Arg Glu Pro Leu Val Lys Ile Trp His Glu Ile Arg Thr Asn
 325 330 335
 10 Gly Pro Lys Lys Gly Gly Gly Ser Lys
 340 345

<210> 409

<211> 236

15 <212> PRT

<213> Homo sapiens

<400> 409

Met Lys Arg Ser Gly Asn Pro Gly Ala Glu Val Thr Asn Ser Ser Val
 1 5 10 15
 Ala Gly Pro Asp Cys Cys Gly Gly Leu Gly Asn Ile Asp Phe Arg Gln
 20 25 30
 Ala Asp Phe Cys Val Met Thr Arg Leu Leu Gly Tyr Val Asp Pro Leu
 35 40 45
 25 Asp Pro Ser Phe Val Ala Ala Val Ile Thr Ile Thr Phe Asn Pro Leu
 50 55 60
 Tyr Trp Asn Val Val Ala Arg Trp Glu His Lys Thr Arg Lys Leu Ser
 65 70 75 80
 Arg Ala Phe Gly Ser Pro Tyr Leu Ala Cys Tyr Ser Leu Ser Ile Thr
 85 90 95
 30 Ile Leu Leu Leu Asn Phe Leu Arg Ser His Cys Phe Thr Gln Ala Met
 100 105 110
 Leu Ser Gln Pro Arg Met Glu Ser Leu Asp Thr Pro Ala Ala Tyr Ser
 115 120 125
 35 Leu Val Leu Ala Leu Leu Gly Leu Gly Val Val Leu Val Leu Ser Ser
 130 135 140
 Phe Phe Ala Leu Gly Phe Ala Gly Thr Phe Leu Gly Asp Tyr Phe Gly
 145 150 155 160
 Ile Leu Lys Glu Ala Arg Val Thr Val Phe Pro Phe Asn Ile Leu Asp
 165 170 175
 40 Asn Pro Met Tyr Trp Gly Ser Thr Ala Asn Tyr Leu Gly Trp Ala Ile
 180 185 190
 Met His Ala Ser Pro Thr Gly Leu Leu Leu Thr Val Leu Val Ala Leu
 195 200 205
 45 Thr Tyr Ile Val Ala Leu Leu Tyr Glu Glu Pro Phe Thr Ala Glu Ile
 210 215 220
 Tyr Arg Gln Lys Ala Ser Gly Ser His Lys Arg Ser
 225 230 235

50 <210> 410

<211> 121

<212> PRT

<213> Homo sapiens

55 <400> 410

Met Asn Thr Glu Ala Glu Gln Gln Leu Leu His His Ala Arg Asn Gly
 1 5 10 15
 Asn Ala Glu Glu Val Arg Gln Leu Leu Glu Thr Met Ala Ser Asn Glu
 20 25 30
 60 Val Ile Ala Asp Ile Asn Cys Lys Gly Arg Ser Lys Ser Asn Leu Gly
 35 40 45
 Trp Thr Pro Leu His Leu Ala Cys Tyr Phe Gly His Arg Gln Val Val
 50 55 60

Gln Asp Leu Leu Lys Ala Gly Ala Glu Val Asn Val Leu Asn Asp Met
 65 70 75 80
 Gly Asp Thr Pro Leu His Arg Ala Ala Phe Thr Gly Arg Lys Val Lys
 85 90 95
 5 Ile Ile Leu Cys Ser Met Phe Val Ser Glu Val Phe Gly Gly Val Val
 100 105 110
 Thr Ile Val Phe Ser Val Ile Thr Ile
 115 120

10 <210> 411
 <211> 170
 <212> PRT
 <213> Homo sapiens

15 <400> 411
 Met Arg Leu Gln Gly Ala Ile Phe Val Leu Leu Pro His Leu Gly Pro
 1 5 10 15
 Ile Leu Val Trp Leu Phe Thr Arg Asp His Met Ser Gly Trp Cys Glu
 20 20 25 30
 Gly Pro Arg Met Leu Ser Trp Cys Pro Phe Tyr Lys Val Leu Leu Leu
 35 40 45
 Val Gln Thr Ala Ile Tyr Ser Val Val Gly Tyr Ala Ser Tyr Leu Val
 50 55 60
 Trp Lys Asp Leu Gly Gly Gly Leu Gly Trp Pro Leu Ala Leu Pro Leu
 25 65 70 75 80
 Gly Leu Tyr Ala Val Gln Leu Thr Ile Ser Trp Thr Val Leu Val Leu
 85 90 95
 Phe Phe Thr Val His Asn Pro Gly Leu Ala Leu Leu His Leu Leu Leu
 100 105 110
 30 Leu Tyr Gly Leu Val Val Ser Thr Ala Leu Ile Trp His Pro Ile Asn
 115 120 125
 Lys Leu Ala Ala Leu Leu Leu Leu Pro Tyr Leu Ala Trp Leu Thr Val
 130 135 140
 Thr Ser Ala Leu Thr Tyr His Leu Trp Arg Asp Ser Leu Cys Pro Val
 35 145 150 155 160
 His Gln Pro Gln Pro Thr Glu Lys Ser Asp
 165 170

<210> 412
 40 <211> 236
 <212> PRT
 <213> Homo sapiens

<400> 412
 45 Met Leu Ser Lys Gly Leu Lys Arg Lys Arg Glu Glu Glu Glu Glu Lys
 1 5 10 15
 Glu Pro Leu Ala Val Asp Ser Trp Trp Leu Asp Pro Gly His Thr Ala
 20 25 30
 Val Ala Gln Ala Pro Pro Ala Val Ala Ser Ser Ser Leu Phe Asp Leu
 35 40 45
 50 Ser Val Leu Lys Leu His His Ser Leu Gln Gln Ser Glu Pro Asp Leu
 50 55 60
 Arg His Leu Val Leu Val Val Asn Thr Leu Arg Arg Ile Gln Ala Ser
 65 70 75 80
 55 Met Ala Pro Ala Ala Ala Leu Pro Pro Val Pro Ser Pro Pro Ala Ala
 85 90 95
 Pro Ser Val Ala Asp Asn Leu Leu Ala Ser Ser Asp Ala Ala Leu Ser
 100 105 110
 Ala Ser Met Ala Ser Leu Leu Glu Asp Leu Ser His Ile Glu Gly Leu
 60 115 120 125
 Ser Gln Ala Pro Gln Pro Leu Ala Asp Glu Gly Pro Pro Gly Arg Ser
 130 135 140
 Ile Gly Gly Ala Ala Pro Ser Leu Gly Ala Leu Asp Leu Leu Gly Pro

145 150 155 160
 Ala Thr Gly Cys Leu Leu Asp Asp Gly Leu Glu Gly Leu Phe Glu Asp
 165 170 175
 5 Ile Asp Thr Ser Met Tyr Asp Asn Glu Leu Trp Ala Pro Ala Ser Glu
 180 185 190
 Gly Leu Lys Pro Gly Pro Glu Asp Gly Pro Gly Lys Glu Glu Ala Pro
 195 200 205
 Glu Leu Asp Glu Ala Glu Leu Asp Tyr Leu Met Asp Val Leu Val Gly
 210 215 220
 10 Thr Gln Ala Leu Glu Arg Pro Pro Gly Pro Gly Arg
 225 230 235

<210> 413

<211> 191

15 <212> PRT

<213> Homo sapiens

<400> 413

Met Lys Gly Leu Tyr Phe Gln Gln Ser Ser Thr Asp Glu Glu Ile Thr
 20 1 5 10 15
 Phe Val Phe Gln Glu Lys Glu Asp Leu Pro Val Thr Glu Asp Asn Phe
 20 25 30
 Val Lys Leu Gln Val Lys Ala Cys Ala Leu Ser Gln Ile Asn Thr Lys
 35 40 45
 25 Leu Leu Ala Glu Met Lys Met Lys Lys Asp Leu Phe Pro Val Gly Arg
 50 55 60
 Glu Ile Ala Gly Ile Val Leu Asp Val Gly Ser Lys Val Ser Phe Phe
 65 70 75 80
 Gln Pro Asp Asp Glu Val Val Gly Ile Leu Pro Leu Asp Ser Glu Asp
 30 85 90 95
 Pro Gly Leu Cys Glu Val Val Arg Val His Glu His Tyr Leu Val His
 100 105 110
 Lys Pro Glu Lys Val Thr Trp Thr Glu Ala Ala Gly Ser Ile Arg Asp
 115 120 125
 35 Gly Val Arg Ala Tyr Thr Ala Leu His Tyr Leu Ser His Leu Ser Pro
 130 135 140
 Gly Lys Ser Val Leu Ile Met Asp Gly Ala Ser Ala Phe Gly Thr Ile
 145 150 155 160
 Ala Ile Gln Leu Ala His His Arg Gly Ala Lys Val Phe Gln Gln His
 40 165 170 175
 Ala Ala Leu Lys Ile Ser Ser Ala Leu Lys Asp Ser Asp Leu Pro
 180 185 190

<210> 414

45 <211> 389

<212> PRT

<213> Homo sapiens

<400> 414

Met Ala Glu Pro Asp Pro Ser His Pro Leu Glu Thr Gln Ala Gly Lys
 50 1 5 10 15
 Val Gln Glu Ala Gln Asp Ser Asp Ser Asp Ser Glu Gly Gly Ala Ala
 20 25 30
 Gly Gly Glu Ala Asp Met Asp Phe Leu Arg Asn Leu Phe Ser Gln Thr
 55 35 40 45
 Leu Ser Leu Gly Ser Gln Lys Glu Arg Leu Leu Asp Glu Leu Thr Leu
 50 55 60
 Glu Gly Val Ala Arg Tyr Met Gln Ser Glu Arg Cys Arg Arg Val Ile
 65 70 75 80
 60 Cys Leu Val Gly Ala Gly Ile Ser Thr Ser Ala Gly Ile Pro Asp Phe
 85 90 95
 Arg Ser Pro Ser Thr Gly Leu Tyr Asp Asn Leu Glu Lys Tyr His Leu
 100 105 110

Pro Tyr Pro Glu Ala Ile Phe Glu Ile Ser Tyr Phe Lys Lys His Pro
 115 120 125
 Glu Pro Phe Phe Ala Leu Ala Lys Glu Leu Tyr Pro Gly Gln Phe Lys
 130 135 140
 5 Pro Thr Ile Cys His Tyr Phe Met Arg Leu Leu Lys Asp Lys Gly Leu
 145 150 155 160
 Leu Leu Arg Cys Tyr Thr Gln Asn Ile Asp Thr Leu Glu Arg Ile Ala
 165 170 175
 10 Gly Leu Glu Gln Glu Asp Leu Val Glu Ala His Gly Thr Phe Tyr Thr
 180 185 190
 Ser His Cys Val Ser Ala Ser Cys Arg His Glu Tyr Pro Leu Ser Trp
 195 200 205
 Met Lys Glu Lys Ile Phe Ser Glu Val Thr Pro Lys Cys Glu Asp Cys
 210 215 220
 15 Gln Ser Leu Val Lys Pro Asp Ile Val Phe Phe Gly Glu Ser Leu Pro
 225 230 235 240
 Ala Arg Phe Phe Ser Cys Met Gln Ser Asp Phe Leu Lys Val Asp Leu
 245 250 255
 20 Leu Leu Val Met Gly Thr Ser Leu Gln Val Gln Pro Phe Ala Ser Leu
 260 265 270
 Ile Ser Lys Ala Pro Leu Ser Thr Pro Arg Leu Leu Ile Asn Lys Glu
 275 280 285
 Lys Ala Gly Gln Ser Asp Pro Phe Leu Gly Met Ile Met Gly Leu Gly
 290 295 300
 25 Gly Gly Met Asp Phe Asp Ser Lys Lys Ala Tyr Arg Asp Val Ala Trp
 305 310 315 320
 Leu Gly Glu Cys Asp Gln Gly Cys Leu Ala Glu Leu Leu Gly
 325 330 335
 30 Trp Lys Lys Glu Leu Glu Asp Leu Val Arg Arg Glu His Ala Ser Ile
 340 345 350
 Asp Ala Gln Ser Gly Ala Gly Val Pro Asn Pro Ser Thr Ser Ala Ser
 355 360 365
 Pro Lys Lys Ser Pro Pro Pro Ala Lys Asp Glu Ala Arg Thr Thr Glu
 370 375 380
 35 Arg Glu Lys Pro Gln
 385

 <210> 415
 <211> 481
 40 <212> PRT
 <213> Homo sapiens

 <400> 415
 Met Ser Leu Asn Leu Pro Glu Ala Ser Leu Leu Ser Arg Ala Ser Trp
 1 5 10 15
 45 Pro Glu Gln Ala Lys Glu Pro Arg Arg Glu Gly His Thr Asp Lys Gln
 20 25 30
 Gln Thr Glu Asp Val Leu Ala Ala Gly Leu Arg Cys Leu Pro His Leu
 35 40 45
 50 Pro Ala Ile Cys Ala Arg Arg Met Ser Pro Ala Phe Arg Ala Met Asp
 50 55 60
 Val Glu Pro Arg Ala Lys Gly Val Leu Leu Glu Pro Phe Val His Gln
 65 70 75 80
 Val Gly Gly His Ser Cys Val Leu Arg Phe Asn Glu Thr Thr Leu Cys
 85 90 95
 55 Lys Pro Leu Val Pro Arg Glu His Gln Phe Tyr Glu Thr Leu Pro Ala
 100 105 110
 Glu Met Arg Lys Phe Thr Pro Gln Tyr Lys Gly Val Val Ser Val Arg
 115 120 125
 60 Phe Glu Glu Asp Glu Asp Arg Asn Leu Cys Leu Ile Ala Tyr Pro Leu
 130 135 140
 Lys Gly Asp His Gly Ile Val Asp Ile Val Asp Asn Ser Asp Cys Glu
 145 150 155 160

Pro Lys Ser Lys Leu Leu Arg Trp Thr Thr Asn Lys Lys His His Val
 165 170 175
 Leu Glu Thr Glu Lys Thr Pro Lys Asp Trp Val Arg Gln His Arg Lys
 180 185 190
 5 Glu Glu Lys Met Lys Ser His Lys Leu Glu Glu Glu Phe Glu Trp Leu
 195 200 205
 Lys Lys Ser Glu Val Leu Tyr Tyr Thr Val Glu Lys Lys Gly Asn Ile
 210 215 220
 10 Ser Ser Gln Leu Lys His Tyr Asn Pro Trp Ser Met Lys Cys His Gln
 225 230 235 240
 Gln Gln Leu Gln Arg Met Lys Glu Asn Ala Lys His Arg Asn Gln Tyr
 245 250 255
 Lys Phe Ile Leu Leu Glu Asn Leu Thr Ser Arg Tyr Glu Val Pro Cys
 260 265 270
 15 Val Leu Asp Leu Lys Met Gly Thr Arg Gln His Gly Asp Asp Ala Ser
 275 280 285
 Glu Glu Lys Ala Ala Asn Gln Ile Arg Lys Cys Gln Gln Ser Thr Ser
 290 295 300
 20 Ala Val Ile Gly Val Arg Val Cys Gly Met Gln Val Tyr Gln Ala Gly
 305 310 315 320
 Ser Gly Gln Leu Met Phe Met Asn Lys Tyr His Gly Arg Lys Leu Ser
 325 330 335
 Val Gln Gly Phe Lys Glu Ala Leu Phe Gln Phe Phe His Asn Gly Arg
 340 345 350
 25 Tyr Leu Arg Arg Glu Leu Leu Gly Pro Val Leu Lys Lys Leu Thr Glu
 355 360 365
 Leu Lys Ala Val Leu Glu Arg Gln Glu Ser Tyr Arg Phe Tyr Ser Ser
 370 375 380
 30 Ser Leu Leu Val Ile Tyr Asp Gly Lys Glu Arg Pro Glu Val Val Leu
 385 390 395 400
 Asp Ser Asp Ala Glu Asp Leu Glu Asp Leu Ser Glu Glu Ser Ala Asp
 405 410 415
 Glu Ser Ala Gly Ala Tyr Ala Tyr Lys Pro Ile Gly Ala Ser Ser Val
 420 425 430
 35 Asp Val Arg Met Ile Asp Phe Ala His Thr Thr Cys Arg Leu Tyr Gly
 435 440 445
 Glu Asp Thr Val Val His Glu Gly Gln Asp Ala Gly Tyr Ile Phe Gly
 450 455 460
 40 Leu Gln Ser Leu Ile Asp Ile Val Thr Glu Ile Ser Glu Glu Ser Gly
 465 470 475 480
 Glu

<210> 416

<211> 354

45 <212> PRT

<213> Homo sapiens

<400> 416

50 Met Ser Ala Gly Gly Gly Arg Ala Phe Ala Trp Gln Val Phe Pro Pro
 1 5 10 15
 Met Pro Thr Cys Arg Val Tyr Gly Thr Val Ala His Gln Asp Gly His
 20 25 30
 Leu Leu Val Leu Gly Gly Cys Gly Arg Ala Gly Leu Pro Leu Asp Thr
 35 40 45
 55 Ala Glu Thr Leu Asp Met Ala Ser His Thr Trp Leu Ala Leu Ala Pro
 50 55 60
 Leu Pro Thr Ala Arg Ala Gly Ala Ala Val Val Leu Gly Lys Gln
 65 70 75 80
 Val Leu Val Val Gly Gly Val Asp Glu Val Gln Ser Pro Val Ala Ala
 85 90 95
 60 Val Glu Ala Phe Leu Met Asp Glu Gly Arg Trp Glu Arg Arg Ala Thr
 100 105 110
 Leu Pro Gln Ala Ala Met Gly Val Ala Thr Val Glu Arg Asp Gly Met

115 120 125
 Val Tyr Ala Leu Gly Gly Met Gly Pro Asp Thr Ala Pro Gln Ala Gln
 130 135 140
 Val Arg Val Tyr Glu Pro Arg Arg Asp Cys Trp Leu Ser Leu Pro Ser
 5 145 150 155 160
 Met Pro Thr Pro Cys Tyr Gly Ala Ser Thr Phe Leu His Gly Asn Lys
 165 170 175
 Ile Tyr Val Leu Gly Gly Arg Gln Gly Lys Leu Pro Val Thr Ala Phe
 180 185 190
 10 Glu Ala Phe Asp Leu Glu Ala Arg Thr Trp Thr Arg His Pro Ser Leu
 195 200 205
 Pro Ser Arg Arg Ala Phe Ala Gly Cys Ala Met Ala Glu Gly Ser Val
 210 215 220
 Phe Ser Leu Gly Gly Leu Gln Gln Pro Gly Pro His Asn Phe Tyr Ser
 15 225 230 235 240
 Arg Pro His Phe Val Asn Thr Val Glu Met Phe Asp Leu Glu His Gly
 245 250 255
 Ser Trp Thr Lys Leu Pro Arg Ser Leu Arg Met Arg Asp Lys Arg Ala
 260 265 270
 20 Asp Phe Val Val Gly Ser Leu Gly Gly His Ile Val Ala Ile Gly Gly
 275 280 285
 Leu Gly Asn Gln Pro Cys Pro Leu Gly Ser Val Glu Ser Phe Ser Leu
 290 295 300
 Ala Arg Arg Arg Trp Glu Ala Leu Pro Ala Met Pro Thr Ala Arg Cys
 25 305 310 315 320
 Ser Cys Ser Ser Leu Gln Ala Gly Pro Arg Leu Phe Val Ile Gly Gly
 325 330 335
 Val Ala Gln Gly Pro Ser Gln Ala Val Glu Ala Leu Cys Leu Arg Asp
 340 345 350
 30 Gly Val

 <210> 417
 <211> 20
 <212> PRT
 35 <213> Homo sapiens

 <400> 417
 Met Lys Gly Leu Tyr Phe Gln Gln Ser Ser Thr Asp Glu Glu Ile Thr
 1 5 10 15
 40 Phe Val Phe Gln
 20

 <210> 418
 <211> 320
 45 <212> PRT
 <213> Homo sapiens

 <400> 418
 Met Lys Gly Leu Tyr Phe Gln Gln Ser Ser Thr Asp Glu Glu Ile Thr
 50 1 5 10 15
 Phe Val Phe Gln Glu Lys Glu Asp Leu Pro Val Thr Glu Asp Asn Phe
 20 25 30
 Val Lys Leu Gln Val Lys Ala Cys Ala Leu Ser Gln Ile Asn Thr Lys
 35 40 45
 55 Leu Leu Ala Glu Met Lys Met Lys Lys Asp Leu Phe Pro Val Gly Arg
 50 55 60
 Glu Ile Ala Gly Ile Val Leu Asp Val Gly Ser Lys Val Ser Phe Phe
 65 70 75 80
 Gln Pro Asp Asp Glu Val Val Gly Ile Leu Pro Leu Asp Ser Glu Asp
 85 90 95
 60 Pro Gly Leu Cys Glu Val Val Arg Val His Glu His Tyr Leu Val His
 100 105 110
 Lys Pro Glu Lys Val Thr Trp Thr Glu Ala Ala Gly Ser Ile Arg Asp

```

      115      120      125
Gly Val Arg Ala Tyr Thr Ala Leu His Tyr Leu Ser His Leu Ser Pro
      130      135      140
5 Gly Lys Ser Val Leu Ile Met Asp Gly Ala Ser Ala Phe Gly Thr Ile
145      150      155      160
Ala Ile Gln Leu Ala His His Arg Gly Ala Lys Val Ile Ser Thr Ala
      165      170      175
Cys Ser Leu Glu Asp Lys Gln Cys Leu Glu Arg Phe Arg Pro Pro Ile
      180      185      190
10 Ala Arg Val Ile Asp Val Ser Asn Gly Lys Val His Val Ala Glu Ser
      195      200      205
Cys Leu Glu Glu Thr Gly Gly Leu Gly Val Asp Ile Val Leu Asp Ala
      210      215      220
Gly Val Arg Leu Tyr Ser Lys Asp Asp Glu Pro Ala Val Lys Leu Gln
15 225      230      235      240
Leu Leu Pro His Lys His Asp Ile Ile Thr Leu Leu Gly Val Gly Gly
      245      250      255
His Trp Val Thr Thr Glu Glu Asn Leu Gln Leu Asp Pro Pro Asp Ser
      260      265      270
20 His Cys Leu Phe Leu Lys Gly Ala Thr Leu Ala Phe Leu Asn Asp Glu
      275      280      285
Val Trp Asn Leu Ser Asn Val Gln Gln Gly Lys Tyr Leu Tyr Leu Lys
      290      295      300
Gly Cys Asp Gly Glu Val Ile Asn Trp Cys Phe Gln Thr Ser Val Gly
25 305      310      315      320

```

<210> 419

<211> 159

<212> PRT

30 <213> Homo sapiens

<400> 419

```

Met Glu Lys Leu Arg Arg Val Leu Ser Gly Gln Asp Asp Glu Glu Gln
1      5      10      15
35 Gly Leu Thr Ala Gln Val Leu Asp Ala Ser Ser Leu Ser Phe Asn Thr
      20      25      30
Arg Leu Lys Trp Phe Ala Ile Cys Phe Val Cys Gly Val Phe Phe Ser
      35      40      45
Ile Leu Gly Thr Gly Leu Leu Trp Leu Pro Gly Gly Ile Lys Leu Phe
40 50      55      60
Ala Val Phe Tyr Thr Leu Gly Asn Leu Ala Ala Leu Ala Ser Thr Cys
      65      70      75      80
Phe Leu Met Gly Pro Val Lys Gln Leu Lys Lys Met Phe Glu Ala Thr
      85      90      95
45 Arg Leu Leu Ala Thr Ile Val Met Leu Leu Cys Phe Ile Phe Thr Leu
      100      105      110
Cys Ala Ala Leu Trp Trp His Lys Lys Gly Leu Ala Val Leu Phe Cys
      115      120      125
Ile Leu Gln Phe Leu Ser Met Thr Trp Tyr Ser Leu Ser Tyr Ile Pro
50 130      135      140
Tyr Ala Arg Asp Ala Val Ile Lys Cys Cys Ser Ser Leu Leu Ser
145      150      155

```

<210> 420

55 <211> 183

<212> PRT

<213> Homo sapiens

<400> 420

```

60 Met Glu Gln Arg Leu Ala Glu Phe Arg Ala Ala Arg Lys Arg Ala Gly
1      5      10      15
Leu Ala Ala Gln Pro Pro Ala Ala Ser Gln Gly Ala Gln Thr Pro Gly
      20      25      30

```

Glu Lys Ala Glu Ala Ala Ala Thr Leu Lys Ala Ala Pro Gly Trp Leu
 35 40 45
 Lys Arg Phe Leu Val Trp Lys Pro Arg Pro Ala Ser Ala Arg Ala Gln
 50 55 60
 5 Pro Gly Leu Val Gln Glu Ala Ala Gln Pro Gln Gly Ser Thr Ser Glu
 65 70 75 80
 Thr Pro Trp Asn Thr Ala Ile Pro Leu Pro Ser Cys Trp Asp Gln Ser
 85 90 95
 10 Phe Leu Thr Asn Ile Thr Phe Leu Lys Val Leu Leu Trp Leu Val Leu
 100 105 110
 Leu Gly Leu Phe Val Glu Leu Glu Phe Gly Leu Ala Tyr Phe Val Leu
 115 120 125
 Ser Leu Phe Tyr Trp Met Tyr Val Gly Thr Arg Gly Pro Glu Glu Lys
 130 135 140
 15 Lys Glu Gly Glu Lys Ser Ala Tyr Ser Val Phe Asn Pro Gly Cys Glu
 145 150 155 160
 Ala Ile Gln Gly Thr Leu Thr Ala Glu Gln Leu Glu Arg Glu Leu Gln
 165 170 175
 Leu Arg Pro Leu Ala Gly Arg
 20 180

<210> 421

<211> 143

<212> PRT

25 <213> Homo sapiens

<400> 421

Met Ala Ala Pro Arg Arg Gly Arg Gly Ser Ser Thr Val Leu Ser Ser
 1 5 10 15
 30 Val Pro Leu Gln Met Leu Phe Tyr Leu Ser Gly Thr Tyr Tyr Ala Leu
 20 25 30
 Tyr Phe Leu Ala Thr Leu Leu Met Ile Thr Tyr Lys Ser Gln Val Phe
 35 40 45
 35 Ser Tyr Pro His Arg Tyr Leu Val Leu Asp Leu Ala Leu Leu Phe Leu
 50 55 60
 Met Gly Ile Leu Glu Ala Val Arg Leu Tyr Leu Gly Thr Arg Gly Asn
 65 70 75 80
 Leu Thr Glu Ala Glu Arg Pro Leu Ala Ala Ser Leu Ala Leu Thr Ala
 85 90 95
 40 Gly Thr Ala Leu Leu Ser Ala His Phe Leu Leu Trp Gln Ala Leu Val
 100 105 110
 Leu Trp Ala Asp Trp Ala Leu Ser Ala Thr Leu Leu Ala Leu His Gly
 115 120 125
 Leu Glu Ala Val Leu Gln Val Val Ala Ile Ala Ala Phe Thr Arg
 45 130 135 140

<210> 422

<211> 73

<212> PRT

50 <213> Homo sapiens

<400> 422

Met Ser Gly Val Pro Ala Glu Met Thr Gly Ala Val Glu Ala Phe Leu
 1 5 10 15
 55 Pro Val Val Ser Ser Ser Arg Arg Leu Pro Arg Phe Val His Met Val
 20 25 30
 Ala Gly Val Ser Ser Lys Gln Glu Arg Ala Arg Ser Asn Thr Glu Ala
 35 40 45
 Leu Phe Lys Leu Cys Phe His His Ile Cys Gln Cys Leu Thr Asp Glu
 50 55 60
 60 His Lys Phe His Gly Gln Val Gln Phe
 65 70

<210> 423
 <211> 142
 <212> PRT
 <213> Homo sapiens

5

<400> 423
 Met Pro Pro Phe Gly Gly His Pro Leu Ser Gln Glu Glu Asp Gly Ser
 1 5 10 15
 Gln Arg Cys Cys Cys Leu Ser Ser Leu Arg Ser Val Asp Asp Ser Asn
 10 20 25 30
 Gly Glu Thr Val Val Ile Met Ala Leu Phe Leu Ala Val Ser Tyr His
 35 40 45
 His Lys Thr Gln Ser Lys Arg Trp Pro Gly Leu Thr Pro Pro His Ser
 50 55 60
 15 Ser Leu Leu Cys Arg Pro Leu Gln Leu Ser Phe Leu Val Ile Gln Ser
 65 70 75 80
 Val Arg Met Arg Ala Cys Gly Cys Asp Ser Gly His Cys Arg Ile Leu
 85 90 95
 Gly Arg Tyr Ser Leu Leu Gly Trp Ser Gln Gly His Arg Ala Arg Gly
 100 105 110
 20 Arg Gly Gly Val Ser Leu Arg Asp Asn Thr Phe Phe Gln Glu Ala Ser
 115 120 125
 Glu Gly Gln Gly Gln Trp Leu Met Pro Val Ile Pro Ala Phe
 130 135 140

25

<210> 424
 <211> 149
 <212> PRT
 <213> Homo sapiens

30

<400> 424
 Met Leu Ser Ile Leu Lys Pro Arg Arg Ser Gln Glu Trp Arg Thr Ala
 1 5 10 15
 Leu Arg Arg Tyr Cys Cys Pro Thr Asp Leu Gln Ala Pro Arg Ser Pro
 15 20 25 30
 Val Pro Pro Ile Arg Lys Val Gly Ile Ser Asp Val Ile Val His Ala
 35 40 45
 Asn Leu Ala Thr Ser Leu Lys Lys Asn Thr Cys Asn Cys Gln Ala Asp
 50 55 60
 40 Leu Leu Ser Trp Arg Ser Trp Val Asn Gly Ile Ser Cys His Cys Pro
 65 70 75 80
 Asn Leu Arg Pro Leu Ser Lys Ser Ile Phe Arg Asp Ser Thr Ser Leu
 85 90 95
 Cys Ser Leu Ser Gln Gln Arg Leu Cys Pro Leu His Ser Lys Pro Glu
 100 105 110
 45 Ala Cys Trp Gly Leu Phe Val Ser Val His Ala His Phe Arg Val Gln
 115 120 125
 Ala Gly Gly Arg Gly Asn Arg Val Gly Lys Lys Thr Arg Val Ser Arg
 130 135 140
 50 Asn Asp Glu Thr Leu
 145

<210> 425
 <211> 75
 55 <212> PRT
 <213> Homo sapiens

<400> 425
 Met Tyr Leu Pro Pro Asn Arg Ser Glu Leu Cys Asn Phe Ala Leu Ser
 1 5 10 15
 Leu Asn Leu Tyr Gly Lys Gly Phe Phe Ser Leu Val Glu Lys His Asn
 20 25 30
 Ser Arg Asp Leu Glu Asp Arg Ala Ser Ser Gly Pro Ser Leu Ser Ser

35 40 45
 Pro Ser His Pro Asp Trp Gly Tyr Ile Val Leu Ile Leu Val Ala Thr
 50 55 60
 Leu Gly Glu Leu Asp Thr Gln Val Gly Gly His
 5 65 70 75

<210> 426

<211> 168

<212> PRT

10 <213> Homo sapiens

<400> 426

Met Arg Leu Thr Glu Lys Ser Glu Gly Glu Gln Gln Leu Lys Pro Asn
 1 5 10 15
 15 Asn Ser Asn Ala Pro Asn Glu Asp Gln Glu Glu Glu Ile Gln Gln Ser
 20 25 30
 Glu Gln His Thr Pro Ala Arg Gln Arg Thr Gln Arg Ala Asp Thr Gln
 35 40 45
 20 Pro Ser Arg Cys Arg Leu Pro Ser Arg Arg Thr Pro Thr Thr Ser Ser
 50 55 60
 Asp Arg Thr Ile Asn Leu Leu Glu Val Leu Pro Trp Pro Thr Glu Trp
 65 70 75 80
 Ile Phe Asn Pro Tyr Arg Leu Pro Ala Leu Phe Glu Leu Tyr Pro Glu
 85 90 95
 25 Phe Leu Leu Val Phe Lys Glu Ala Phe His Asp Ile Ser His Cys Leu
 100 105 110
 Lys Ala Gln Met Glu Lys Ile Gly Leu Pro Ile Ile Leu His Leu Phe
 115 120 125
 30 Ala Leu Ser Thr Leu Tyr Phe Tyr Lys Phe Phe Leu Pro Thr Ile Leu
 130 135 140
 Ser Leu Ser Phe Phe Ile Leu Leu Val Leu Leu Leu Leu Phe Ile
 145 150 155 160
 Ile Val Phe Ile Leu Ile Phe Phe
 165

35

<210> 427

<211> 160

<212> PRT

<213> Homo sapiens

40

<400> 427

Met Pro Arg Ser Ser Arg Ser Pro Gly Asp Pro Gly Ala Leu Leu Glu
 1 5 10 15
 45 Asp Val Ala His Asn Pro Arg Pro Arg Ile Ala Gln Arg Gly Arg
 20 25 30
 Asn Thr Ser Arg Met Ala Glu Asp Thr Ser Pro Asn Met Asn Asp Asn
 35 40 45
 Ile Leu Leu Pro Val Arg Asn Asn Asp Gln Ala Leu Gly Leu Thr Gln
 50 55 60
 Cys Met Leu Gly Cys Val Ser Trp Phe Thr Cys Phe Ala Cys Ser Leu
 65 70 75 80
 Arg Thr Gln Ala Gln Val Leu Phe Asn Thr Cys Arg Asp Arg Val
 85 90 95
 Ser Pro Cys Cys Pro Gly Trp Ser Gln Thr Pro Val Ile Leu Pro Pro
 55 100 105 110
 Gln Pro Ser Glu Val Leu Gly Leu Gln Met Gln Ala Ala Val Pro Glu
 115 120 125
 Ala His Gly Glu Asp Arg His Ser Ala Pro Leu Cys Phe Arg Cys Val
 130 135 140
 60 Pro Gly Pro Cys Pro Val Pro Gly Gly Gly Ile Pro Gly Pro Trp His
 145 150 155 160

<210> 428

<211> 94
 <212> PRT
 <213> Homo sapiens

5 <400> 428
 Met Asn Lys Glu Ile Asp Ser Leu Asn Leu Ala Tyr Ser Phe Pro Phe
 1 5 10 15
 Leu Leu Pro Ala Phe Leu Asp Thr Pro Trp Thr Asp Pro Phe Pro Ser
 20 25 30
 10 Gly Phe Met Val Arg Ser Arg Val Leu Leu Ile Gln Leu Leu Ser Arg
 35 40 45
 Pro Arg Ser Ser Gln Glu Ser Arg Gly His Ser Leu Pro Cys Ser Pro
 50 55 60
 Ser Ala Leu His Lys Pro Gly Gly Ile Cys Pro Ala Ala Leu Gly Arg
 15 65 70 75 80
 Ser His Leu Leu Val Trp Glu Gln Pro Ser Leu Arg Asp Ser
 85 90

<210> 429
 20 <211> 95
 <212> PRT
 <213> Homo sapiens

<400> 429
 25 Met Lys Ala Ser Gly Pro Asp Leu Ser Asp Gly Leu His Cys Pro Ser
 1 5 10 15
 Leu Ile Arg His Leu Arg Thr Phe Ser Ala Ala Ala Leu Ala Pro
 20 25 30
 Arg Tyr Pro Thr Arg Leu Pro Ser Ser Leu Leu Leu Trp His Leu Cys
 30 35 40 45
 Gln Cys Leu His Leu Leu Tyr Ala Val Ser Thr Ser Cys Asn Ser His
 50 55 60
 Gly Lys Arg Ser Ala Ala Trp Ala Met Thr Arg Thr Glu Asp Thr Asp
 65 70 75 80
 35 Ala Leu Thr Asp Ser Phe Asp Asp Ser Phe Ile Ser Ser Ala Asp
 85 90 95

<210> 430
 <211> 99
 40 <212> PRT
 <213> Homo sapiens

<400> 430
 45 Met Lys Lys Lys Glu Glu Thr Thr Leu Ser Glu Met Glu Pro Val Glu
 1 5 10 15
 Pro Gln Tyr Gln Leu Val Asn Ala Glu Ser Thr Ser Pro Phe Leu His
 20 25 30
 Cys Leu Arg Glu Val Ile Gly Glu Tyr Ser Val His Glu Phe Ser Leu
 35 40 45
 50 Leu Gly Lys Thr Glu Ser Gln Gly Ile Gly Leu Trp Ile Ala Leu Val
 50 55 60
 Val Phe Leu Ser Phe Leu Ile Phe Ser Thr Ser Phe Tyr Ile Ser Asn
 65 70 75 80
 Ala Glu Gln Pro Phe Phe Lys Glu Pro Pro Thr Glu Ala Ala Lys Glu
 55 85 90 95
 Leu Ser Leu

<210> 431
 <211> 122
 60 <212> PRT
 <213> Homo sapiens

<400> 431

Ile Arg Ala Thr Met Val Ala Arg Val Trp Ser Leu Met Arg Phe Leu
 1 5 10 15
 Ile Lys Gly Ser Val Ala Gly Gly Ala Val Tyr Leu Val Tyr Asp Gln
 20 25 30
 5 Glu Leu Leu Gly Pro Ser Asp Lys Ser Gln Ala Ala Leu Gln Lys Ala
 35 40 45
 Gly Glu Val Val Pro Pro Ala Met Tyr Gln Phe Ser Gln Tyr Val Cys
 50 55 60
 Gln Gln Thr Gly Leu Gln Ile Pro Gln Leu Pro Ala Pro Pro Lys Ile
 10 65 70 75 80
 Tyr Phe Pro Ile Arg Asp Ser Trp Asn Ala Gly Ile Met Thr Val Met
 85 90 95
 Ser Ala Leu Ser Val Ala Pro Ser Lys Ala Arg Glu Tyr Ser Lys Glu
 100 105 110
 15 Gly Trp Glu Tyr Val Lys Ala Arg Thr Lys
 115 120

<210> 432

<211> 118

20 <212> PRT

<213> Homo sapiens

<400> 432

Met Gln Pro Ser Leu Leu Arg Ser Tyr Arg Leu Lys Ala Gln Leu Ser
 25 1 5 10 15
 Leu Ser Ser Thr Val Pro Arg Arg Ile Thr Asp Lys Pro Ala Thr Lys
 20 25 30
 Ser Trp Glu Gly Gly Arg Arg Glu Leu Cys Pro Arg Val Leu Phe Thr
 35 40 45
 30 Gln Leu Leu Leu Trp Val Trp Pro Gly Asp Pro Gly Pro Glu Leu Gln
 50 55 60
 Glu Thr Gly Phe Pro Gly Pro Pro Arg Pro Ala His Leu Lys Thr Asp
 65 70 75 80
 Arg Ala Ile Met Val Gly Val Lys Gly Ile Glu Glu Lys Ser Gly Ile
 35 85 90 95
 Gly Ala Gly Val Cys Arg Val Ser Val Glu Lys Leu Ala Ser Thr Gln
 100 105 110
 Glu Arg Thr Ser Ser Leu
 115

40

<210> 433

<211> 49

<212> PRT

<213> Homo sapiens

45

<400> 433

Met Glu Leu Glu Ala Met Ser Arg Tyr Thr Ser Pro Val Asn Pro Pro
 1 5 10 15
 Val Phe Pro His Leu Thr Val Val Leu Leu Ala Ile Gly Met Phe Phe
 20 25 30
 Thr Ala Trp Phe Phe Val Tyr Pro Phe Thr Glu Gln Pro Glu Asp Gln
 35 40 45
 His

55 <210> 434

<211> 89

<212> PRT

<213> Homo sapiens

60 <400> 434

Met Leu Ala Leu Phe His Phe His Leu Pro Pro Trp Asp Asp Ala Val
 1 5 10 15
 Arg Arg Pro Ser Val Asp Ala Ser Pro Ser Thr Leu Asn Phe Pro Asp

20 25 30
 Ala Glu Leu Tyr Ala Ser Ile Phe Leu Cys Cys Met Ala Pro Gly Glu
 35 40 45
 Ile Leu Ile Ser Phe Leu Thr Leu Val Gln Ile Ala His Ala Asn Gly
 5 50 55 60
 Arg Gly Cys Asn Thr Pro Ala Cys Gly Ala Ala Ala Cys Val Trp His
 65 70 75 80
 Glu Asn Ser Gln Glu Glu Arg Lys Tyr
 85
 10
 <210> 435
 <211> 87
 <212> PRT
 <213> Homo sapiens
 15
 <400> 435
 Met Ser Gln Gln His Arg Arg Lys Arg Pro Ser Ser Glu Arg Lys Ser
 1 5 10 15
 Thr Arg Lys Met Asp Thr Trp Gln Ser Leu Lys Val Lys Glu Val Phe
 20 20 25 30
 Cys Lys His Asn Ser Ser Tyr Glu Cys Leu Leu Tyr Lys Glu Val Glu
 35 40 45
 Ala Arg Gln Val Ser Lys Thr Ala Thr Asp Gly Ser Tyr Leu Leu Val
 50 55 60
 25 Phe Thr Ser Tyr Val Ile Ser Ser Pro Val Trp Thr Gly Pro Gly Asp
 65 70 75 80
 Leu Leu Pro Val Asn Arg Ile
 85
 30
 <210> 436
 <211> 45
 <212> PRT
 <213> Homo sapiens
 35
 <400> 436
 Met Pro Arg Ser Ser Arg Ser Pro Gly Asp Pro Gly Ala Leu Leu Glu
 1 5 10 15
 Asp Gly Pro Gln Ser Gln Thr Pro Glu Asp Cys Pro Ala Arg Pro Glu
 20 25 30
 40 His Gln Gln Asp Gly Arg Gly His Leu Pro Lys His Glu
 35 40 45
 <210> 437
 <211> 65
 45
 <212> PRT
 <213> Homo sapiens
 <400> 437
 Met Ala Tyr Leu Asp Asp Lys Gly Ser Leu Leu Ala Ile His Ser His
 1 5 10 15
 Ala Arg Gln His Ser His Glu Thr Asn Gln Val His Gln Trp Leu Pro
 20 25 30
 Arg Asn Thr Phe Ala Phe Leu Ile Lys Glu Asp Arg Cys Ser Cys Arg
 35 40 45
 55 Ser Thr Cys Ala Ser Phe Ser Phe Ser Ser Ser Phe Ser Phe Leu Ile
 50 55 60
 Ser
 65
 60
 <210> 438
 <211> 112
 <212> PRT
 <213> Homo sapiens

<400> 438
 Met Arg Lys Lys Cys Lys Cys Phe Thr Ile Lys Lys Thr Asn Thr Tyr
 1 5 10 15
 5 Glu Glu Ser Asn Ala Gly Asn Glu Gly Gln Lys Glu Ala Ile Ser Ile
 20 25 30
 Cys Ile Cys Arg Arg Asp Gly Leu Leu Pro Leu Trp Val Thr Arg Leu
 35 40 45
 Ser Asp Leu Val Phe Ser Lys Glu Lys Ala His Gly Met Ile Pro Leu
 10 50 55 60
 Leu Gly Ser His Arg Glu Lys Lys Thr Ser Lys Glu Met Lys Thr Ser
 65 70 75 80
 Ser Arg Asn Leu Arg Tyr Phe Ile Val Cys Arg Asp Ala Ser Ser Tyr
 85 90 95
 15 Thr Pro Gln Ser Leu Ile Ser Gly Tyr Ile Gly Pro Cys Gln His Gln
 100 105 110

<210> 439
 <211> 110
 20 <212> PRT
 <213> Homo sapiens

<400> 439
 Met Val Phe Gly Ala Met Val Leu Leu Val Gly Leu Glu Glu Leu Thr
 1 5 10 15
 Asn Ile Arg Asn Val Glu Arg Leu Lys Lys Asp Leu Arg Ala Ser Tyr
 20 25 30
 Cys Leu Ile Asp Ser Phe Leu Gly Asp Ser Glu Leu Ile Gly Asp Leu
 35 40 45
 30 Thr Gln Cys Val Asp Cys Val Ile Pro Pro Glu Gly Ser Leu Leu Gln
 50 55 60
 Ile Ser Ser Tyr Leu Tyr Leu Asn Thr Ala Leu Val Asp Leu Pro Gly
 65 70 75 80
 Val Ala Ala Ser Gln Ala Cys Asp Ser Gln Gln Val Thr Trp Leu Leu
 35 85 90 95
 Tyr Val Ala Asn Gly Ala Tyr Ser Ala Cys Asn Arg Pro Gly
 100 105 110

<210> 440
 40 <211> 121
 <212> PRT
 <213> Homo sapiens

<400> 440
 45 Thr Ser Ser Ser Gly Ala Glu Val Thr Met Ala Ala Ala Leu Ala Arg
 1 5 10 15
 Leu Gly Leu Arg Pro Val Lys Gln Val Arg Val Gln Phe Cys Pro Phe
 20 25 30
 Glu Lys Asn Val Glu Ser Thr Arg Thr Phe Leu Gln Thr Val Ser Ser
 35 40 45
 50 Glu Lys Val Arg Ser Thr Asn Leu Asn Cys Ser Val Ile Ala Asp Val
 50 55 60
 Arg His Asp Gly Ser Glu Pro Cys Val Asp Val Leu Phe Gly Asp Gly
 65 70 75 80
 55 His Arg Leu Ile Met Arg Gly Ala His Leu Thr Ala Leu Glu Met Leu
 85 90 95
 Thr Ala Phe Ala Ser His Ile Arg Ala Arg Asp Ala Ala Gly Ser Gly
 100 105 110
 Asp Lys Pro Gly Ala Asp Thr Gly Arg
 60 115 120

<210> 441
 <211> 99

<212> PRT

<213> Homo sapiens

<400> 441

5 Met Leu Ala Arg Ala Thr Phe Arg Ala Ala Ser Ala Pro Thr Leu Val
 1 5 10 15
 Ala Arg Arg Gly Phe Gln Ser Thr Arg Ala Gln Met Ala Ser Pro Tyr
 20 25 30
 His Tyr Pro Glu Gly Pro Arg Ser Asn Leu Pro Phe Asp Pro Leu Lys
 35 40 45
 Lys Gly Phe Ala Phe Lys Tyr Trp Gly Phe Met Gly Thr Gly Phe Ala
 50 55 60
 Leu Pro Phe Leu Leu Ala Val Trp Gln Thr Glu Gln Ala Val Asn Ala
 65 70 75 80
 15 Leu Arg His Gly Val Asp Met Arg Ile Gly Ile Pro Gly Asn Thr Ala
 85 90 95
 Phe Val Asp

<210> 442

20 <211> 183

<212> PRT

<213> Homo sapiens

<400> 442

25 Arg Glu Gly Ala Arg Ala Arg Pro Ser Pro Thr Met Ser Asp Glu Ala
 1 5 10 15
 Ser Ala Ile Thr Ser Tyr Glu Lys Phe Leu Thr Pro Glu Glu Pro Phe
 20 25 30
 Pro Leu Leu Gly Pro Pro Arg Gly Val Gly Thr Cys Pro Ser Glu Glu
 35 40 45
 Pro Gly Cys Leu Asp Ile Ser Asp Phe Gly Cys Gln Leu Ser Ser Cys
 50 55 60
 His Arg Thr Asp Pro Leu His Arg Phe His Thr Asn Arg Trp Asn Leu
 65 70 75 80
 35 Thr Ser Cys Gly Thr Ser Val Ala Ser Ser Glu Gly Ser Glu Glu Leu
 85 90 95
 Phe Ser Ser Val Ser Val Gly Asp Gln Asp Asp Cys Tyr Ser Leu Leu
 100 105 110
 Asp Asp Gln Asp Phe Thr Ser Phe Asp Leu Phe Pro Glu Gly Ser Val
 115 120 125
 Cys Ser Asp Val Ser Ser Ser Ile Ser Thr Tyr Trp Asp Trp Ser Asp
 130 135 140
 Ser Glu Phe Glu Trp Gln Leu Pro Gly Ser Asp Ile Ala Ser Gly Ser
 145 150 155 160
 45 Asp Val Leu Ser Asp Val Ile Pro Ser Ile Pro Ser Ser Pro Cys Leu
 165 170 175
 Leu Pro Lys Lys Lys Lys Lys
 180

50 <210> 443

<211> 94

<212> PRT

<213> Homo sapiens

55 <400> 443

Met Ser Asp Glu Ala Ser Ala Ile Thr Ser Tyr Glu Lys Phe Leu Thr
 1 5 10 15
 Pro Glu Glu Pro Phe Pro Leu Leu Gly Pro Pro Arg Gly Val Gly Thr
 20 25 30
 60 Cys Pro Ser Glu Glu Pro Gly Cys Leu Asp Ile Ser Asp Phe Gly Cys
 35 40 45
 Gln Leu Ser Ser Cys His Arg Thr Asp Pro Leu His Arg Phe His Thr
 50 55 60

Asn Arg Trp Asn Leu Thr Ser Cys Gly Thr Ser Val Ala Ser Ser Glu
 65 70 75 80
 Gly Ser Glu Glu Leu Phe Ser Ser Val Cys Trp Arg Ser Arg
 85 90

5
 <210> 444
 <211> 105
 <212> PRT
 <213> Homo sapiens

10
 <400> 444
 Ile Gly Pro Arg Ala Pro Ser Pro Ser Phe Ser Val Arg Asp Val Glu
 1 5 10 15
 Leu Ser Asp Pro Ala Arg Glu Arg Gly Glu Met Pro Val Ala Val Gly
 20 25 30
 Pro Tyr Gly Gln Ser Gln Pro Ser Cys Phe Asp Arg Val Lys Met Gly
 35 40 45
 Phe Val Met Gly Cys Ala Val Gly Met Ala Ala Gly Ala Leu Phe Gly
 50 55 60
 20 Thr Phe Ser Cys Leu Arg Ile Gly Met Arg Gly Arg Glu Leu Met Gly
 65 70 75 80
 Gly Ile Gly Lys Thr Met Met Gln Ser Gly Gly Thr Phe Gly Thr Phe
 85 90 95
 Met Ala Ile Gly Met Gly Ile Arg Cys
 25 100 105

<210> 445
 <211> 163
 <212> PRT
 30 <213> Homo sapiens

<400> 445
 Met Pro Arg Ser Ser Arg Ser Pro Gly Asp Pro Gly Ala Leu Leu Glu
 1 5 10 15
 35 Asp Val Ala His Asn Pro Arg Pro Arg Ile Ala Gln Arg Gly Arg
 20 25 30
 Asn Thr Ser Arg Met Ala Glu Asp Thr Ser Pro Asn Met Asn Asp Asn
 35 40 45
 Ile Leu Leu Pro Val Arg Asn Asn Asp Gln Ala Leu Gly Leu Thr Gln
 40 50 55 60
 Cys Met Leu Gly Cys Val Ser Trp Phe Thr Cys Phe Ala Cys Ser Leu
 65 70 75 80
 Arg Thr Gln Ala Gln Gln Val Leu Phe Asn Thr Cys Arg Cys Lys Leu
 85 90 95
 45 Leu Cys Gln Lys Leu Met Glu Lys Thr Gly Ile Leu Leu Leu Cys Ala
 100 105 110
 Phe Gly Val Ser Gln Gly Pro Ala Gln Ser Gln Val Glu Val Ser Leu
 115 120 125
 Gly Pro Gly Thr Asp Tyr Arg Thr Leu Gly Lys Thr Leu His Cys His
 50 130 135 140
 Val Thr Gln Phe Pro His Leu Pro Asp Gly Cys Cys Cys Glu Asn Tyr
 145 150 155 160
 Glu Met Lys

55 <210> 446
 <211> 128
 <212> PRT
 <213> Homo sapiens

60 <400> 446
 Met Glu Asp Lys Glu Ile Pro Ile Lys Ser Glu Pro Leu Pro Lys Pro
 1 5 10 15
 Pro Ala Ser Ala Pro Pro Ser Ile Leu Val Lys Pro Glu Asn Ser Arg

```

                20                25                30
      Asn Gly Ile Glu Lys Gln Val Lys Thr Val Arg Phe Gln Asn Tyr Ser
                35                40                45
      Pro Pro Pro Thr Lys His Tyr Thr Ser His Pro Thr Ser Gly Lys Pro
5      50                55                60
      Glu Gln Pro Ala Thr Leu Lys Ala Ser Gln Pro Glu Ala Ala Ser Leu
      65                70                75                80
      Gly Pro Glu Met Thr Val Leu Phe Ala His Arg Ser Gly Cys His Ser
                85                90                95
10     Gly Gln Gln Thr Asp Leu Arg Arg Lys Ser Ala Leu Ala Lys Ala Thr
                100                105                110
      Thr Leu Val Ser Thr Ala Ser Gly Thr Gln Thr Val Phe Pro Ser Lys
                115                120                125

```

15 <210> 447
 <211> 96
 <212> PRT
 <213> Homo sapiens

```

20 <400> 447
      Met Leu Thr Arg Val Glu Glu Gln Lys Lys Met Val Lys Ala Cys Arg
      1                5                10                15
      Tyr Arg Cys Ser Ala Cys His Leu Lys Tyr Ser Pro Gln Arg Gln Lys
                20                25                30
25     Glu Arg Lys Leu Ser Leu Lys Arg Gly Arg Thr Ser Gln Gln Asn Met
                35                40                45
      Ser Met Phe Trp Leu Lys Lys Leu Leu Glu Ser Gly Leu Phe Cys Ala
                50                55                60
      Met Cys Ser Pro Arg Ala Ser Thr Lys Lys Gly Phe Trp Cys Arg Pro
30     65                70                75                80
      Lys Thr Thr Ile Ile Ile Ile Asp Tyr Ser Ser Pro Arg Gln Cys Leu
                85                90                95

```

<210> 448
 35 <211> 160
 <212> PRT
 <213> Homo sapiens

<220>
 40 <221> UNSURE
 <222> 114
 <223> Xaa = Glu,Val

<220>
 45 <221> UNSURE
 <222> 113
 <223> Xaa = His,Gln

<220>
 50 <221> UNSURE
 <222> 115
 <223> Xaa = Ile,Val

```

<400> 448
55     Met Gly Lys Ile Ala Leu Gln Leu Lys Ala Thr Leu Glu Asn Ile Thr
      1                5                10                15
      Asn Leu Arg Pro Val Gly Glu Asp Phe Arg Trp Tyr Leu Lys Met Lys
                20                25                30
      Cys Gly Asn Cys Gly Glu Ile Ser Asp Lys Trp Gln Tyr Ile Arg Leu
60     35                40                45
      Met Asp Ser Val Ala Leu Lys Gly Gly Arg Gly Ser Ala Ser Met Val
                50                55                60
      Gln Lys Cys Lys Leu Cys Ala Arg Glu Asn Ser Ile Glu Ile Leu Ser

```

	65				70				75				80
	Ser	Thr	Ile	Lys	Pro	Tyr	Asn	Ala	Glu	Asp	Asn	Glu	Thr
					85				90				95
	Ile	Val	Glu	Phe	Glu	Cys	Arg	Gly	Leu	Glu	Pro	Val	Thr
5				100					105				110
	Xaa	Xaa	Xaa	Leu	Leu	Leu	Lys	Val	Trp	Ser	Gln	Gly	Val
				115					120				125
	Thr	Leu	Ile	Cys	Arg	Arg	Arg	Thr	Gly	Thr	Asp	Tyr	Ala
				130					135				140
10	Gln	Glu	Ser	Val	Gly	Ile	Tyr	Glu	Val	Thr	His	Gln	Cys
	145					150					155		160

<210> 449

<211> 117

15 <212> PRT

<213> Homo sapiens

<400> 449

20	Met	Asp	Ser	Leu	Ala	Ala	Gly	Glu	Leu	Asn	Ala	Ser	His	Gln	Pro	Trp
1				5						10					15	
	Val	Pro	Glu	Phe	Val	Ala	Tyr	Trp	Arg	Lys	Thr	His	Gln	Asp	His	Leu
				20					25					30		
	Cys	Ser	Leu	His	Ser	Arg	Ala	Phe	Gly	Leu	Leu	Asp	Ala	Arg	Val	Thr
				35					40					45		
25	Trp	Ala	Leu	Arg	Arg	Ala	Pro	Glu	Pro	Val	Pro	Gly	Lys	Asp	Arg	Leu
				50					55					60		
	Leu	Leu	Ala	Ala	Phe	Pro	Ala	Glu	Ala	Ser	Pro	Val	Asp	Thr	Ala	Ser
	65					70						75				80
	Val	Ser	Val	Tyr	Gly	Arg	Ala	Pro	Arg	Tyr	Met	His	Lys	Gly	Val	Lys
30				85						90					95	
	Lys	Cys	Val	Cys	Thr	Pro	Val	Ser	Lys	Asn	Ser	Thr	Ala	Trp	Leu	Leu
				100					105					110		
	Leu	Gly	Gly	Ile	Ser											
				115												

35

<210> 450

<211> 335

<212> PRT

<213> Homo sapiens

40

<400> 450

	Met	Cys	Cys	Gln	Val	Cys	Glu	Ala	Val	Arg	Ser	Gly	Asn	Glu	Glu	Val
1				5						10					15	
	Leu	Ala	Asp	Val	Arg	Thr	Ile	Val	Asn	Gln	Ile	Ser	Tyr	Thr	Pro	Gln
45				20					25					30		
	Asp	Pro	Arg	Asp	Leu	Cys	Gly	Arg	Ile	Leu	Thr	Thr	Cys	Tyr	Met	Ala
				35					40					45		
	Ser	Lys	Asn	Ser	Ser	Gln	Glu	Thr	Cys	Thr	Arg	Ala	Arg	Glu	Leu	Ala
				50					55					60		
50	Gln	Gln	Ile	Gly	Ser	His	His	Ile	Ser	Leu	Asn	Ile	Asp	Pro	Ala	Val
	65					70						75				80
	Lys	Ala	Val	Met	Gly	Ile	Phe	Ser	Leu	Val	Thr	Gly	Lys	Ser	Pro	Leu
				85						90					95	
	Phe	Ala	Ala	His	Gly	Gly	Ser	Ser	Arg	Glu	Asn	Leu	Ala	Leu	Gln	Asn
55				100					105					110		
	Val	Gln	Ala	Arg	Ile	Arg	Met	Val	Leu	Ala	Tyr	Leu	Phe	Ala	Gln	Leu
				115					120					125		
	Ser	Leu	Trp	Ser	Arg	Gly	Val	His	Gly	Gly	Leu	Leu	Val	Leu	Gly	Ser
				130					135					140		
60	Ala	Asn	Val	Asp	Glu	Ser	Leu	Leu	Gly	Tyr	Leu	Thr	Lys	Tyr	Asp	Cys
	145					150						155				160
	Ser	Ser	Ala	Asp	Ile	Asn	Pro	Ile	Gly	Gly	Ile	Ser	Lys	Thr	Asp	Leu
					165					170					175	

Arg Ala Phe Val Gln Phe Cys Ile Gln Arg Phe Gln Leu Pro Ala Leu
 180 185 190
 Gln Ser Ile Leu Leu Ala Pro Ala Thr Ala Glu Leu Glu Pro Leu Ala
 195 200 205
 5 Asp Gly Gln Val Ser Gln Thr Asp Glu Glu Asp Met Gly Met Thr Tyr
 210 215 220
 Ala Glu Leu Ser Val Tyr Gly Lys Leu Arg Lys Val Ala Lys Met Gly
 225 230 235 240
 10 Pro Tyr Ser Met Phe Cys Lys Leu Leu Gly Met Trp Arg His Ile Cys
 245 250 255
 Thr Pro Arg Gln Val Ala Asp Lys Val Lys Arg Phe Phe Ser Lys Tyr
 260 265 270
 Ser Met Asn Arg His Lys Met Thr Thr Leu Thr Pro Ala Tyr His Ala
 275 280 285
 15 Glu Asn Tyr Ser Pro Glu Asp Asn Arg Phe Asp Leu Arg Pro Phe Leu
 290 295 300
 Tyr Asn Thr Ser Trp Pro Trp Gln Phe Arg Cys Ile Glu Asn Gln Val
 305 310 315 320
 20 Leu Gln Leu Glu Arg Ala Glu Pro Gln Ser Leu Asp Gly Val Asp
 325 330 335

<210> 451

<211> 86

<212> PRT

25 <213> Homo sapiens

<220>

<221> UNSURE

<222> 76

30 <223> Xaa = Lys, Asn

<400> 451

Met Cys Trp Val Ile Asn His Ala Ile Leu Pro Arg Met Arg Met His
 1 5 10 15
 35 Ser Lys Arg Gln Thr Ile Thr Arg His Ser Ala Ser Leu Ser Phe His
 20 25 30
 Ala Leu Pro Arg Ser Ala Phe Leu Gln Leu Cys Leu Leu Arg Gln Ile
 35 40 45
 40 His Gln Ile Pro Cys Leu Ser Ile Phe Ser Ser Thr Leu Arg Ala Gln
 50 55 60
 Thr His Asp Ser Gly Ile Gly Cys Thr Thr Ala Xaa Pro Gly Gly Arg
 65 70 75 80
 Arg Gln Glu Gln Leu Arg
 85

45

<210> 452

<211> 93

<212> PRT

<213> Homo sapiens

50

<400> 452

Met Lys Ile Ala Leu Cys Gln Arg Glu Leu Pro Ser Pro Arg Ser Cys
 1 5 10 15
 55 Leu Leu Ser Arg Asp Val Thr Gly Val Ile Cys Thr Arg Met Pro Arg
 20 25 30
 Leu Ala Ile Cys Ser Lys Thr Ala Gln Lys Ala Leu Pro Cys Ile Pro
 35 40 45
 Leu Leu His Thr Ser Pro Leu Cys Leu Gln Leu Leu Ser Ala Gly Leu
 50 55 60
 60 His Ile Tyr Ala Thr Leu Cys Lys Ser Cys Ala Ser Arg Asn His Lys
 65 70 75 80
 Asn Ile Phe Leu His Leu Leu His Ser Leu Ser Ala Ala
 85 90

<210> 453

<211> 108

<212> PRT

5 <213> Homo sapiens

<400> 453

Met Ala Val Arg Ala Ser Phe Glu Asn Asn Cys Glu Ile Gly Cys Phe
 1 5 10 15
 10 Ala Lys Leu Thr Asn Thr Tyr Cys Leu Val Ala Ile Gly Gly Ser Glu
 20 25 30
 Asn Phe Tyr Ser Val Phe Glu Gly Glu Leu Ser Asp Thr Ile Pro Val
 35 40 45
 Val His Ala Ser Ile Ala Gly Cys Arg Ile Ile Gly Arg Met Cys Val
 15 50 55 60
 Gly Asp Arg Arg Asn Ser Gly Arg Cys Ala Gln Gly Gly Ser Leu Gln
 65 70 75 80
 Thr Asp Ser Gly Arg Pro Gly Ala Ser Arg Lys Leu Leu Cys Leu Gln
 85 90 95
 20 Gln Ser Gly Arg Ala Gly Ala Ser Gln Asp Phe Asn
 100 105

<210> 454

<211> 277

25 <212> PRT

<213> Homo sapiens

<400> 454

Met Ser Leu Cys Glu Asp Met Leu Leu Cys Asn Tyr Arg Lys Cys Arg
 1 5 10 15
 30 Ile Lys Leu Ser Gly Tyr Ala Trp Val Thr Ala Cys Ser His Ile Phe
 20 25 30
 Cys Asp Gln His Gly Ser Gly Glu Phe Ser Arg Ser Pro Ala Ile Cys
 35 35 40 45
 Pro Ala Cys Asn Ser Thr Leu Ser Gly Lys Leu Asp Ile Val Arg Thr
 50 55 60
 Glu Leu Ser Pro Ser Glu Glu Tyr Lys Ala Met Val Leu Ala Gly Leu
 65 70 75 80
 Arg Pro Glu Ile Val Leu Asp Ile Ser Ser Arg Ala Leu Ala Phe Trp
 40 85 90 95
 Thr Tyr Gln Val His Gln Glu Arg Leu Tyr Gln Glu Tyr Asn Phe Ser
 100 105 110
 Lys Ala Glu Gly His Leu Lys Gln Met Glu Lys Ile Tyr Thr Gln Gln
 115 120 125
 45 Ile Gln Ser Lys Asp Val Glu Leu Thr Ser Met Lys Gly Glu Val Thr
 130 135 140
 Ser Met Lys Lys Val Leu Glu Glu Tyr Lys Lys Lys Phe Ser Asp Ile
 145 150 155 160
 Ser Glu Lys Leu Met Glu Arg Asn Arg Gln Tyr Gln Lys Leu Gln Gly
 50 165 170 175
 Leu Tyr Asp Ser Leu Arg Leu Arg Asn Ile Thr Ile Ala Asn His Glu
 180 185 190
 Gly Thr Leu Glu Pro Ser Met Ile Ala Gln Ser Gly Val Leu Gly Phe
 195 200 205
 55 Pro Leu Gly Asn Asn Ser Lys Phe Pro Leu Asp Asn Thr Pro Val Arg
 210 215 220
 Asn Arg Gly Asp Gly Asp Gly Asp Phe Gln Phe Arg Pro Phe Phe Ala
 225 230 235 240
 Gly Ser Pro Thr Ala Pro Glu Pro Ser Asn Ser Phe Phe Ser Phe Val
 60 245 250 255
 Ser Pro Ser Arg Glu Leu Glu Gln Gln Val Ser Ser Arg Ala Phe
 260 265 270
 Lys Val Lys Arg Ile

275

<210> 455

<211> 173

5 <212> PRT

<213> Homo sapiens

<400> 455

10 Met Leu Val Met Tyr Leu Leu Ala Ala Leu Phe Gly Tyr Leu Thr Phe
 1 5 10 15
 Tyr Gly Glu Val Glu Asp Glu Leu Leu His Ala Tyr Ser Lys Val Tyr
 20 25 30
 Thr Leu Asp Ile Pro Leu Leu Met Val Arg Leu Ala Val Leu Val Ala
 35 40 45
 15 Val Thr Leu Thr Val Pro Ile Val Leu Phe Pro Ile Arg Thr Ser Val
 50 55 60
 Ile Thr Leu Leu Phe Pro Lys Arg Pro Phe Ser Trp Ile Arg His Phe
 65 70 75 80
 Leu Ile Ala Ala Val Leu Ile Ala Leu Asn Asn Val Leu Val Ile Leu
 20 85 90 95
 Val Pro Thr Ile Lys Tyr Ile Phe Gly Phe Ile Gly Ala Ser Ser Ala
 100 105 110
 Thr Met Leu Ile Phe Ile Leu Pro Ala Val Phe Tyr Leu Lys Leu Val
 115 120 125
 25 Lys Lys Glu Thr Phe Arg Ser Pro Gln Lys Val Gly Ala Leu Ile Phe
 130 135 140
 Leu Val Val Gly Ile Phe Phe Met Ile Gly Ser Met Ala Leu Ile Ile
 145 150 155 160
 Ile Asp Trp Ile Tyr Asp Pro Pro Asn Ser Lys His His
 30 165 170

<210> 456

<211> 370

<212> PRT

35 <213> Homo sapiens

<400> 456

Met Ser Ala Ser Ala Ala Thr Gly Val Phe Val Leu Ser Leu Ser Ala
 1 5 10 15
 40 Ile Pro Val Thr Tyr Val Phe Asn His Leu Ala Ala Gln His Asp Ser
 20 25 30
 Trp Thr Ile Val Gly Val Ala Ala Leu Ile Leu Phe Leu Val Ala Leu
 35 40 45
 Leu Ala Arg Val Leu Val Lys Arg Lys Pro Pro Arg Asp Pro Leu Phe
 45 50 55 60
 Tyr Val Tyr Ala Val Phe Gly Phe Thr Ser Val Val Asn Leu Ile Ile
 65 70 75 80
 Gly Leu Glu Gln Asp Gly Ile Ile Asp Gly Phe Met Thr His Tyr Leu
 85 90 95
 50 Arg Glu Gly Glu Pro Tyr Leu Asn Thr Ala Tyr Gly His Met Ile Cys
 100 105 110
 Tyr Trp Asp Gly Ser Ala His Tyr Leu Met Tyr Leu Val Met Val Ala
 115 120 125
 Ala Ile Ala Trp Glu Glu Thr Tyr Arg Thr Ile Gly Leu Tyr Trp Val
 55 130 135 140
 Gly Ser Ile Ile Met Ser Val Val Val Phe Val Pro Gly Asn Ile Val
 145 150 155 160
 Gly Lys Tyr Gly Thr Arg Ile Cys Pro Ala Phe Phe Leu Ser Ile Pro
 165 170 175
 60 Tyr Thr Cys Leu Pro Val Trp Ala Gly Phe Arg Ile Tyr Asn Gln Pro
 180 185 190
 Ser Glu Asn Tyr Asn Tyr Pro Ser Lys Val Ile Gln Glu Ala Gln Ala
 195 200 205

Lys Asp Leu Leu Arg Arg Pro Phe Asp Leu Met Leu Val Val Cys Leu
 210 215 220
 Leu Leu Ala Thr Gly Phe Cys Leu Phe Arg Gly Leu Ile Ala Leu Asp
 225 230 235 240
 5 Cys Pro Ser Glu Leu Cys Arg Leu Tyr Thr Gln Phe Gln Glu Pro Tyr
 245 250 255
 Leu Lys Asp Pro Ala Ala Tyr Pro Lys Ile Gln Met Leu Ala Tyr Met
 260 265 270
 10 Phe Tyr Ser Val Pro Tyr Phe Val Thr Ala Leu Tyr Gly Leu Val Val
 275 280 285
 Pro Gly Cys Ser Trp Met Pro Asp Ile Thr Leu Ile His Ala Gly Gly
 290 295 300
 Leu Ala Gln Ala Gln Phe Ser His Ile Gly Ala Ser Leu His Ala Arg
 305 310 315 320
 15 Thr Ala Tyr Val Tyr Arg Val Pro Glu Glu Ala Lys Ile Leu Phe Leu
 325 330 335
 Ala Leu Asn Ile Ala Tyr Gly Val Leu Pro Gln Leu Leu Ala Tyr Arg
 340 345 350
 20 Cys Ile Tyr Lys Pro Glu Phe Phe Ile Lys Thr Lys Ala Glu Glu Lys
 355 360 365
 Val Glu
 370

 <210> 457
 25 <211> 393
 <212> PRT
 <213> Homo sapiens

 <400> 457
 30 Met Thr Tyr Arg Trp Gly Thr Leu Leu Met Lys Arg Lys Phe Glu Glu
 1 5 10 15
 Pro Arg Pro Gly Phe His Gly Val Leu Gly Ile Asn Ser Ile Thr Gly
 20 25 30
 35 Lys Glu Glu Pro Leu Tyr Pro Ser Tyr Lys Arg Gln Leu Arg Ile Tyr
 35 40 45
 Leu Val Ser Leu Pro Phe Val Cys Leu Cys Leu Tyr Phe Ser Leu Tyr
 50 55 60
 Val Met Met Ile Tyr Phe Asp Met Glu Val Trp Ala Leu Gly Leu His
 65 70 75 80
 40 Glu Asn Ser Gly Ser Glu Trp Thr Ser Val Leu Leu Tyr Val Pro Ser
 85 90 95
 Ile Ile Tyr Ala Ile Val Ile Glu Ile Met Asn Arg Leu Tyr Arg Tyr
 100 105 110
 45 Ala Ala Glu Phe Leu Thr Ser Trp Glu Asn His Arg Leu Glu Ser Ala
 115 120 125
 Tyr Gln Asn His Leu Ile Leu Lys Val Leu Val Phe Asn Phe Leu Asn
 130 135 140
 Cys Phe Ala Ser Leu Phe Tyr Ile Ala Phe Val Leu Lys Asp Met Lys
 145 150 155 160
 50 Leu Leu Arg Gln Ser Leu Ala Thr Leu Leu Ile Thr Ser Gln Ile Leu
 165 170 175
 Asn Gln Ile Met Glu Ser Phe Leu Pro Tyr Trp Leu Gln Arg Lys His
 180 185 190
 Gly Val Arg Val Lys Arg Lys Val Gln Ala Leu Lys Ala Asp Ile Asp
 195 200 205
 55 Ala Thr Leu Tyr Glu Gln Val Ile Leu Glu Lys Glu Met Gly Thr Tyr
 210 215 220
 Leu Gly Thr Phe Asp Asp Tyr Leu Glu Leu Phe Leu Gln Phe Gly Tyr
 225 230 235 240
 60 Val Ser Leu Phe Ser Cys Val Tyr Pro Leu Ala Ala Ala Phe Ala Val
 245 250 255
 Leu Asn Asn Phe Thr Glu Val Asn Ser Asp Ala Leu Lys Met Cys Arg
 260 265 270

Val Phe Lys Arg Pro Phe Ser Glu Pro Ser Ala Asn Ile Gly Val Trp
 275 280 285
 Gln Leu Ala Phe Glu Thr Met Ser Val Ile Ser Val Val Thr Asn Cys
 290 295 300
 5 Ala Leu Ile Gly Met Ser Pro Gln Val Asn Ala Val Phe Pro Glu Ser
 305 310 315 320
 Lys Ala Asp Leu Ile Leu Ile Val Val Ala Val Glu His Ala Leu Leu
 325 330 335
 10 Ala Leu Lys Phe Ile Leu Ala Phe Ala Ile Pro Asp Lys Pro Arg His
 340 345 350
 Ile Gln Met Lys Leu Ala Arg Leu Glu Phe Glu Ser Leu Glu Ala Leu
 355 360 365
 Lys Gln Gln Gln Met Lys Leu Val Thr Glu Asn Leu Lys Glu Glu Pro
 370 375 380
 15 Met Glu Ser Gly Lys Glu Lys Ala Thr
 385 390

<210> 458

<211> 116

20 <212> PRT

<213> Homo sapiens

<400> 458

25 Met Val Gly Gly Glu Ala Ala Ala Ala Val Glu Glu Leu Val Ser Gly
 1 5 10 15
 Val Arg Gln Ala Ala Asp Phe Ala Glu Gln Phe Arg Ser Tyr Ser Glu
 20 25 30
 Ser Glu Lys Gln Trp Lys Ala Arg Met Glu Phe Ile Leu Arg His Leu
 35 40 45
 30 Pro Asp Tyr Arg Asp Pro Pro Asp Gly Ser Gly Arg Leu Asp Gln Leu
 50 55 60
 Leu Ser Leu Ser Met Val Trp Ala Asn His Leu Phe Leu Gly Cys Ser
 65 70 75 80
 Tyr Asn Lys Asp Leu Leu Asp Lys Val Met Glu Met Ala Asp Gly Ile
 85 90 95
 35 Glu Val Glu Asp Leu Pro Gln Phe Thr Arg Ser Glu Leu Met Lys
 100 105 110
 Lys His Gln Ser
 115

40

<210> 459

<211> 163

<212> PRT

<213> Homo sapiens

45

<400> 459

Met Glu His Tyr Arg Lys Ala Gly Ser Val Glu Leu Pro Ala Pro Ser
 1 5 10 15
 Pro Met Pro Gln Leu Pro Pro Asp Thr Leu Glu Met Arg Val Arg Asp
 20 25 30
 Gly Ser Lys Ile Arg Asn Leu Leu Gly Leu Ala Leu Gly Arg Leu Glu
 35 40 45
 Gly Gly Ser Ala Arg His Val Val Phe Ser Gly Ser Gly Arg Ala Ala
 50 55 60
 55 Gly Lys Ala Val Ser Cys Ala Glu Ile Val Lys Arg Arg Val Pro Gly
 65 70 75 80
 Leu His Gln Leu Thr Lys Leu Arg Phe Leu Gln Thr Glu Asp Ser Trp
 85 90 95
 Val Pro Ala Ser Pro Asp Thr Gly Leu Asp Pro Leu Thr Val Arg Arg
 100 105 110
 60 His Val Pro Ala Val Trp Val Leu Ser Arg Asp Pro Leu Asp Pro
 115 120 125
 Asn Glu Cys Gly Tyr Gln Pro Pro Gly Ala Pro Pro Gly Leu Gly Ser

130 135 140
 Met Pro Ser Ser Ser Cys Gly Pro Arg Ser Arg Arg Arg Ala Arg Asp
 145 150 155 160
 Thr Arg Ser
 5
 <210> 460
 <211> 230
 <212> PRT
 <213> Homo sapiens
 10
 <400> 460
 Met Val Val Phe Gly Tyr Glu Ala Gly Thr Lys Pro Arg Asp Ser Gly
 1 5 10 15
 Val Val Pro Val Gly Thr Glu Glu Ala Pro Lys Val Phe Lys Met Ala
 15 20 25 30
 Ala Ser Met His Gly Gln Pro Ser Pro Ser Leu Glu Asp Ala Lys Leu
 35 40 45
 Arg Arg Pro Met Val Ile Glu Ile Ile Glu Lys Asn Phe Asp Tyr Leu
 50 55 60
 20 Arg Lys Glu Met Thr Gln Asn Ile Tyr Gln Met Ala Thr Phe Gly Thr
 65 70 75 80
 Thr Ala Gly Phe Ser Gly Ile Phe Ser Asn Phe Leu Phe Arg Arg Cys
 85 90 95
 Phe Lys Val Lys His Asp Ala Leu Lys Thr Tyr Ala Ser Leu Ala Thr
 100 105 110
 25 Leu Pro Phe Leu Ser Thr Val Val Thr Asp Lys Leu Phe Val Ile Asp
 115 120 125
 Ala Leu Tyr Ser Asp Asn Ile Ser Lys Glu Asn Cys Val Phe Arg Ser
 130 135 140
 30 Ser Leu Ile Gly Ile Val Cys Gly Val Phe Tyr Pro Ser Ser Leu Ala
 145 150 155 160
 Phe Thr Lys Asn Gly Arg Leu Ala Thr Lys Tyr His Thr Val Pro Leu
 165 170 175
 Pro Pro Lys Gly Arg Val Leu Ile His Trp Met Thr Leu Cys Gln Thr
 180 185 190
 35 Gln Met Lys Leu Met Ala Ile Pro Leu Val Phe Gln Ile Met Phe Gly
 195 200 205
 Ile Leu Asn Gly Leu Tyr His Tyr Ala Val Phe Glu Glu Thr Leu Glu
 210 215 220
 40 Lys Thr Ile His Glu Glu
 225 230
 <210> 461
 <211> 101
 45 <212> PRT
 <213> Homo sapiens
 <220>
 <221> UNSURE
 50 <222> 95
 <223> Xaa = Cys, Trp
 <400> 461
 Met Glu Arg Pro Asp Lys Ala Ala Leu Asn Ala Leu Gln Pro Pro Glu
 1 5 10 15
 Phe Arg Asn Glu Ser Ser Leu Ala Ser Thr Leu Lys Thr Leu Leu Phe
 20 25 30
 Phe Thr Ala Leu Met Ile Thr Val Pro Ile Gly Leu Tyr Phe Thr Thr
 35 40 45
 60 Lys Ser Tyr Ile Phe Glu Gly Ala Leu Gly Met Ser Asn Arg Asp Ser
 50 55 60
 Tyr Phe Tyr Ala Ala Ile Val Ala Val Val Ala Val His Val Val Leu
 65 70 75 80

Ala Leu Phe Val Tyr Val Ala Trp Asn Glu Gly Ser Arg Gln Xaa Arg
85 90 95
Glu Gly Lys Gln Asp
100

5
<210> 462
<211> 93
<212> PRT
<213> Homo sapiens

10
<400> 462
Met Asp Ser Leu Arg Lys Met Leu Ile Ser Val Ala Met Leu Gly Ala
1 5 10 15
Gly Ala Gly Val Gly Tyr Ala Leu Leu Val Ile Val Thr Pro Gly Glu
20 25 30
15 Arg Arg Lys Gln Glu Met Leu Lys Glu Met Pro Leu Gln Asp Pro Arg
35 40 45
Ser Arg Glu Glu Ala Ala Arg Thr Gln Gln Leu Leu Ala Thr Leu
50 55 60
20 Gln Glu Ala Ala Thr Thr Gln Glu Asn Val Ala Trp Arg Lys Asn Trp
65 70 75 80
Met Val Gly Gly Glu Gly Gly Ala Gly Gly Arg Ser Pro
85 90

25
<210> 463
<211> 133
<212> PRT
<213> Homo sapiens

30
<400> 463
Met Gly His Gly Asp Glu Ile Val Leu Ala Asp Leu Asn Phe Pro Ala
1 5 10 15
Ser Ser Ile Cys Gln Cys Gly Pro Met Glu Ile Arg Ala Asp Gly Leu
20 25 30
35 Gly Ile Pro Gln Leu Leu Glu Ala Val Leu Lys Leu Leu Pro Leu Asp
35 40 45
Thr Tyr Val Glu Ser Pro Ala Ala Val Met Glu Leu Val Pro Ser Asp
50 55 60
Lys Glu Arg Gly Leu Gln Thr Pro Val Trp Thr Glu Tyr Glu Ser Ile
40 65 70 75 80
Leu Arg Arg Ala Gly Cys Val Arg Ala Leu Ala Lys Ile Glu Arg Phe
85 90 95
Glu Phe Tyr Glu Arg Ala Lys Lys Ala Phe Ala Val Val Ala Thr Gly
100 105 110
45 Glu Thr Ala Leu Tyr Gly Asn Leu Ile Leu Arg Lys Gly Val Leu Ala
115 120 125
Leu Asn Pro Leu Leu
130

50
<210> 464
<211> 95
<212> PRT
<213> Homo sapiens

55
<400> 464
Met Gly His Gly Asp Glu Ile Val Leu Ala Asp Leu Asn Phe Pro Ala
1 5 10 15
Ser Ser Ile Cys Gln Cys Gly Pro Met Glu Ile Arg Ala Asp Gly Leu
20 25 30
60 Gly Ile Pro Gln Leu Leu Glu Ala Val Leu Ala Ala Ala Pro Gly His
35 40 45
Leu Cys Gly Glu Ser Gly Cys Ser His Gly Ala Gly Ala Gln Arg Gln
50 55 60

Gly Glu Gly Pro Ala Asp Pro Ser Val Asp Gly Val Arg Val His Pro
 65 70 75 80
 Thr Gln Gly Arg Leu Cys Glu Ser Pro Gly Lys Asp Arg Glu Val
 85 90 95

5

<210> 465
 <211> 93
 <212> PRT
 <213> Homo sapiens

10

<400> 465
 Met Thr Pro Ile Lys Leu Leu Asn Leu Thr Ser Arg Tyr Asn Phe Arg
 1 5 10 15
 Arg Thr Phe Gly Ile Glu Leu Ser Ser Asn Ser Ser Tyr Cys Lys Arg
 20 25 30
 Gly Asn Gly Tyr Arg Ser Arg Val Pro Lys Glu Cys Glu Cys Asn Trp
 35 40 45
 Leu His Leu Glu Ser Asp Thr Leu Lys Lys Leu Pro Ile Ile Ser Pro
 50 55 60
 Ser Trp Thr Cys Arg Ile Ile Leu Phe Leu Tyr Phe Ser Gly Gln Leu
 65 70 75 80
 Leu Gln Leu Ser Leu Ser Cys Leu Gln Leu Ile Lys Leu
 85 90

25

<210> 466
 <211> 500
 <212> PRT
 <213> Homo sapiens

30

<400> 466
 Met Glu Val Ser Thr Asn Pro Ser Ser Asn Ile Asp Pro Gly Asn Tyr
 1 5 10 15
 Val Glu Met Asn Asp Ser Ile Thr His Leu Pro Ser Lys Val Val Ile
 20 25 30
 Gln Asp Ile Thr Met Glu Leu His Cys Pro Leu Cys Asn Asp Trp Phe
 35 40 45
 Arg Asp Pro Leu Met Leu Ser Cys Gly His Asn Phe Cys Glu Ala Cys
 50 55 60
 Ile Gln Asp Phe Trp Arg Leu Gln Ala Lys Glu Thr Phe Cys Pro Glu
 65 70 75 80
 Cys Lys Met Leu Cys Gln Tyr Asn Asn Cys Thr Phe Asn Pro Val Leu
 85 90 95
 Asp Lys Leu Val Glu Lys Ile Lys Lys Leu Pro Leu Leu Lys Gly His
 100 105 110
 Pro Gln Cys Pro Glu His Gly Glu Asn Leu Lys Leu Phe Ser Lys Pro
 115 120 125
 Asp Gly Lys Leu Ile Cys Phe Gln Cys Lys Asp Ala Arg Leu Ser Val
 130 135 140
 Gly Gln Ser Lys Glu Phe Leu Gln Ile Ser Asp Ala Val His Phe Phe
 145 150 155 160
 Met Glu Glu Leu Ala Ile Gln Gln Gly Gln Leu Glu Thr Thr Leu Lys
 165 170 175
 Glu Leu Gln Thr Leu Arg Asn Met Gln Lys Glu Ala Ile Ala Ala His
 180 185 190
 Lys Glu Asn Lys Leu His Leu Gln Gln His Val Ser Met Glu Phe Leu
 195 200 205
 Lys Leu His Gln Phe Leu His Ser Lys Glu Lys Asp Ile Leu Thr Glu
 210 215 220
 Leu Arg Glu Glu Gly Lys Ala Leu Asn Glu Glu Met Glu Leu Asn Leu
 225 230 235 240
 Ser Gln Leu Gln Glu Gln Cys Leu Leu Ala Lys Asp Met Leu Val Ser
 245 250 255
 Ile Gln Ala Lys Thr Glu Gln Gln Asn Ser Phe Asp Phe Leu Lys Asp

260 265 270
 Ile Thr Thr Leu Leu His Ser Leu Glu Gln Gly Met Lys Val Leu Ala
 275 280 285
 Thr Arg Glu Leu Ile Ser Arg Lys Leu Asn Leu Gly Gln Tyr Lys Gly
 5 290 295 300
 Pro Ile Gln Tyr Met Val Trp Arg Glu Met Gln Asp Thr Leu Cys Pro
 305 310 315 320
 Gly Leu Ser Pro Leu Thr Leu Asp Pro Lys Thr Ala His Pro Asn Leu
 325 330 335
 10 Val Leu Ser Lys Ser Gln Thr Ser Val Trp His Gly Asp Ile Lys Lys
 340 345 350
 Ile Met Pro Asp Asp Pro Glu Arg Phe Asp Ser Ser Val Ala Val Leu
 355 360 365
 Gly Ser Arg Gly Phe Thr Ser Gly Lys Trp Tyr Trp Glu Val Glu Val
 15 370 375 380
 Ala Lys Lys Thr Lys Trp Thr Val Gly Val Val Arg Glu Ser Ile Ile
 385 390 395 400
 Arg Lys Gly Ser Cys Pro Leu Thr Pro Glu Gln Gly Phe Trp Leu Leu
 405 410 415
 20 Arg Leu Arg Asn Gln Thr Asp Leu Lys Ala Leu Asp Leu Pro Ser Phe
 420 425 430
 Ser Leu Thr Leu Thr Asn Asn Leu Asp Lys Val Gly Ile Tyr Leu Asp
 435 440 445
 Tyr Glu Gly Gly Gln Leu Ser Phe Tyr Asn Ala Lys Thr Met Thr His
 25 450 455 460
 Ile Tyr Thr Phe Ser Asn Thr Phe Met Glu Lys Leu Tyr Pro Tyr Phe
 465 470 475 480
 Cys Pro Cys Leu Asn Asp Gly Arg Glu Asn Lys Glu Pro Leu His Ile
 485 490 495
 30 Leu His Pro Gln
 500

<210> 467

<211> 140

35 <212> PRT

<213> Homo sapiens

<400> 467

40 Met Val Leu Thr Lys Pro Leu Gln Arg Asn Gly Ser Met Met Ser Phe
 1 5 10 15
 Glu Asn Val Lys Glu Lys Ser Arg Glu Gly Gly Pro His Ala His Thr
 20 25 30
 Pro Glu Glu Glu Leu Cys Phe Val Val Thr His Tyr Pro Gln Val Gln
 35 40 45
 45 Thr Thr Leu Asn Leu Phe Phe His Ile Phe Lys Val Leu Thr Gln Pro
 50 55 60
 Leu Ser Leu Leu Trp Gly Cys Asp Gln Lys Pro Arg Thr Val Pro Thr
 65 70 75 80
 Leu Gly Asn Gly Ala Trp Asp Thr Cys Gln Gln His Ile Arg Thr Ser
 85 90 95
 50 Ser Trp Thr Ala Asn Thr Leu Val Ile Gln Asn Gln His Ser Arg Glu
 100 105 110
 Ser Thr Val Ser Val Cys Leu Phe Met Leu Ile Arg Met Gln His Ile
 115 120 125
 55 Leu Lys Thr Asp Thr Leu Gln Gln Phe Arg Ile Cys
 130 135 140

<210> 468

<211> 100

60 <212> PRT

<213> Homo sapiens

<400> 468

Met Tyr Met Leu Leu Ser Pro His Arg Leu Arg Glu Gln Ala Gly Val
 1 5 10 15
 Arg Gly Ser Ile Arg Thr Ala Asn Arg Thr Glu Asp Gly Leu Lys Ile
 20 25 30
 5 Arg Glu Ala Glu Ser Leu Pro Gln Ser Asn Thr Ala Asp Phe Lys Cys
 35 40 45
 Leu His Ser Ala Ser Leu Gln Gln Ala Pro Gly Gly Ile Leu Met Gly
 50 55 60
 Pro Ala Ser Ser Pro Trp Thr Leu Ala Val Glu Gly Glu Lys Arg Thr
 10 65 70 75 80
 Ser Ala Pro Pro Leu Arg Glu Ser Leu Met Pro Thr Lys Gly Leu Gly
 85 90 95
 Trp Trp Thr Gln
 100

15

<210> 469
 <211> 119
 <212> PRT
 <213> Homo sapiens

20

<400> 469
 Met Ala Ser Tyr Ser Gly Phe Ser Gly Leu Leu Glu Ile Arg Tyr Gly
 1 5 10 15
 Pro Gly His Arg Ser Cys Leu Pro Gln Phe Ala Phe Phe Pro Gln Pro
 20 25 30
 25 Pro Leu Pro Arg Pro Arg Ile Cys Met Trp Val Leu Ala Glu Leu Leu
 35 40 45
 Glu Leu Gly Cys Pro Glu Gln Ser Leu Arg Asp Ala Ile Thr Leu Asp
 50 55 60
 30 Leu Phe Cys His Ala Leu Ile Phe Cys Arg Gln Gln Gly Phe Ser Leu
 65 70 75 80
 Glu Gln Thr Ser Ala Ala Cys Ala Leu Leu Gln Asp Leu His Lys Ala
 85 90 95
 Cys Ile Gly Glu Arg Gly Gln Leu Pro Gly Leu Ser Pro Arg Glu Lys
 35 100 105 110
 Arg Asn Arg Ala Trp His Lys
 115

<210> 470
 40 <211> 140
 <212> PRT
 <213> Homo sapiens

<400> 470
 45 Met Arg Ser Glu Cys Val Leu Gly Ala Ala Ser Asp Ser Gly Gln Glu
 1 5 10 15
 Ala Pro Arg Asp Thr Trp Phe Leu Gln Gly Trp Lys Ala Ser Arg Arg
 20 25 30
 Phe Leu Ile Lys Gly Ser Val Ala Gly Gly Ala Val Tyr Leu Val Tyr
 35 40 45
 50 Asp Gln Glu Leu Leu Gly Pro Ser Asp Lys Ser Gln Ala Ala Leu Gln
 50 55 60
 Lys Ala Gly Glu Val Val Pro Pro Ala Met Tyr Gln Phe Ser Gln Tyr
 65 70 75 80
 55 Val Cys Gln Gln Thr Gly Leu Gln Ile Pro Gln Leu Pro Ala Pro Pro
 85 90 95
 Lys Ile Tyr Phe Pro Ile Arg Asp Ser Trp Asn Ala Gly Ile Met Thr
 100 105 110
 Val Met Ser Ala Leu Ser Val Ala Pro Ser Lys Ala Arg Glu Tyr Ser
 60 115 120 125
 Lys Glu Gly Trp Glu Tyr Val Lys Ala Arg Thr Lys
 130 135 140

<210> 471
 <211> 109
 <212> PRT
 <213> Homo sapiens

5

<400> 471
 Met Phe His Leu Arg Thr Cys Ala Ala Lys Leu Arg Pro Leu Thr Ala
 1 5 10 15
 Ser Gln Thr Val Lys Thr Phe Ser Gln Asn Arg Pro Ala Ala Arg
 10 20 25 30
 Thr Phe Gln Gln Ile Arg Cys Tyr Ser Ala Pro Val Ala Ala Glu Pro
 35 40 45
 Phe Leu Ser Gly Thr Ser Ser Asn Tyr Val Glu Glu Met Tyr Cys Ala
 50 55 60
 Trp Leu Glu Asn Pro Lys Ser Val His Lys Thr Gly Ser His Cys Cys
 15 65 70 75 80
 Pro Gly Trp Ser Ala Val Ala Gly Ser Arg Leu Ala Ala Thr Ser Asp
 85 90 95
 Ser Trp Val Gln Val Ile Leu Met Pro Gln Pro Pro Glu
 20 100 105

<210> 472
 <211> 100
 <212> PRT
 25 <213> Homo sapiens

<400> 472
 Met Phe His Leu Arg Thr Cys Ala Ala Lys Leu Arg Pro Leu Thr Ala
 1 5 10 15
 30 Ser Gln Thr Val Lys Thr Phe Ser Gln Asn Arg Pro Ala Ala Ala Arg
 20 25 30
 Thr Phe Gln Gln Ile Arg Ala Ile Leu His Leu Leu Leu Leu Ser Pro
 35 40 45
 Phe Ser Val Gly Leu Val Arg Thr Met Trp Arg Arg Cys Thr Val Leu
 35 50 55 60
 Gly Trp Lys Thr Pro Lys Val Tyr Ile Arg Gln Gly Pro Thr Val Val
 65 70 75 80
 Gln Ala Gly Val Gln Trp Arg Asp Leu Gly Leu Leu Gln Pro Pro Thr
 85 90 95
 40 Pro Gly Phe Lys
 100

<210> 473
 <211> 141
 45 <212> PRT
 <213> Homo sapiens

<400> 473
 Met Ala Pro Lys Val Phe Arg Gln Tyr Trp Asp Ile Pro Asp Gly Thr
 1 5 10 15
 Asp Cys His Arg Lys Ala Tyr Ser Thr Thr Ser Ile Ala Ser Val Ala
 20 25 30
 Gly Leu Thr Ala Ala Ala Tyr Arg Val Thr Leu Asn Pro Pro Gly Thr
 35 40 45
 55 Phe Leu Glu Gly Val Ala Lys Val Gly Gln Tyr Thr Phe Thr Ala Ala
 50 55 60
 Ala Val Gly Ala Val Phe Gly Leu Thr Thr Cys Ile Ser Ala His Val
 65 70 75 80
 Arg Glu Lys Pro Asp Asp Pro Leu Asn Tyr Phe Leu Gly Gly Cys Ala
 85 90 95
 60 Gly Gly Leu Thr Leu Gly Ala Arg Thr His Asn Tyr Gly Ile Gly Ala
 100 105 110
 Ala Ala Cys Val Tyr Phe Gly Ile Ala Ala Ser Leu Val Lys Met Gly

115 120 125
 Arg Leu Glu Gly Trp Glu Val Phe Ala Lys Pro Lys Val
 130 135 140

5 <210> 474
 <211> 134
 <212> PRT
 <213> Homo sapiens

10 <400> 474
 Met Ala Thr His Pro Asp Gly Phe Arg Leu Glu Gly Pro Leu Ala Ala
 1 5 10 15
 Ala His Ser Pro Gly Pro Cys Thr Val Leu Tyr Glu Gly Pro Val Arg
 20 25 30
 15 Gly Leu Cys Pro Phe Ala Pro Arg Asn Ser Asn Thr Met Ala Ala Ala
 35 40 45
 Ala Leu Ala Ala Pro Ser Leu Gly Phe Asp Gly Val Ile Gly Val Leu
 50 55 60
 Val Ala Asp Thr Ser Leu Thr Asp Met His Val Val Asp Val Glu Leu
 20 65 70 75 80
 Ser Gly Pro Arg Gly Pro Thr Gly Arg Ser Phe Ala Val His Thr Arg
 85 90 95
 Arg Glu Asn Pro Ala Glu Pro Gly Ala Val Thr Gly Ser Ala Thr Val
 100 105 110
 25 Thr Ala Phe Trp Arg Ser Leu Leu Ala Cys Cys Gln Leu Pro Ser Arg
 115 120 125
 Pro Gly Ile His Leu Cys
 130

30 <210> 475
 <211> 134
 <212> PRT
 <213> Homo sapiens

35 <400> 475
 Met Ala Thr His Pro Asp Gly Phe Arg Leu Glu Gly Pro Leu Ala Ala
 1 5 10 15
 Ala His Ser Pro Gly Pro Cys Thr Val Leu Tyr Glu Gly Pro Val Arg
 20 25 30
 40 Gly Leu Cys Pro Phe Ala Pro Arg Asn Ser Asn Thr Met Ser Ala Ala
 35 40 45
 Ala Leu Ala Ala Pro Ser Leu Gly Phe Asp Gly Val Ile Gly Val Leu
 50 55 60
 Val Ala Asp Thr Ser Leu Thr Asp Met His Val Val Asp Val Glu Leu
 45 65 70 75 80
 Ser Gly Pro Arg Gly Pro Thr Cys Arg Ser Phe Ala Val His Thr Arg
 85 90 95
 Arg Glu Asn Pro Ala Glu Pro Gly Ala Val Thr Gly Ser Ala Thr Val
 100 105 110
 50 Thr Ala Phe Trp Arg Ser Leu Leu Ala Cys Cys Gln Leu Pro Ser Arg
 115 120 125
 Pro Gly Ile His Leu Cys
 130

55 <210> 476
 <211> 85
 <212> PRT
 <213> Homo sapiens

60 <400> 476
 Met Leu Lys Val Glu Ala Thr Gly Ser Pro Glu Glu Gly Trp Ala Gly
 1 5 10 15
 Gly Glu Pro Arg Thr Gly Ala Pro Ala Asn Ser Pro Ser Cys Pro Gln

20 25 30
 Glu Met Pro Leu Gln Asp Pro Arg Ser Arg Glu Glu Ala Ala Arg Thr
 35 40 45
 Gln Gln Leu Leu Leu Ala Thr Leu Gln Glu Ala Ala Thr Thr Gln Glu
 5 50 55 60
 Asn Val Ala Trp Arg Lys Asn Trp Met Val Gly Gly Glu Gly Gly Ala
 65 70 75 80
 Ser Gly Arg Ser Pro
 85
 10
 <210> 477
 <211> 116
 <212> PRT
 <213> Homo sapiens
 15
 <400> 477
 Met Gly Arg Pro Trp Met Val Met Ile Leu Glu Ser Lys Ser Glu Glu
 1 5 10 15
 Lys Met Trp Tyr Gly Val Phe Leu Trp Ala Leu Val Ser Ser Leu Phe
 20 20 25 30
 Phe His Val Pro Ala Gly Leu Leu Ala Leu Phe Thr Leu Arg His His
 35 40 45
 Lys Tyr Gly Arg Phe Met Ser Val Ser Ile Leu Leu Met Gly Ile Val
 50 55 60
 25 Gly Pro Ile Thr Ala Gly Ile Leu Thr Ser Ala Ala Ile Ala Gly Val
 65 70 75 80
 Tyr Arg Ala Ala Gly Lys Glu Met Ile Pro Phe Glu Ala Leu Thr Leu
 85 90 95
 Gly Thr Gly Gln Thr Phe Cys Val Leu Val Val Ser Phe Leu Arg Ile
 30 100 105 110
 Leu Ala Thr Leu
 115
 <210> 478
 35 <211> 104
 <212> PRT
 <213> Homo sapiens
 <400> 478
 40 Met Asn Arg Tyr Cys Gly Lys Ile Phe Val Ser Val Met Val Lys Leu
 1 5 10 15
 Gln Lys Asn Lys Leu Thr Ser Phe Pro Arg Gln Pro Leu Leu Thr Phe
 20 25 30
 Phe Glu Tyr Leu Glu Lys Val Leu Cys Ser Gly Leu Phe Ser His Ser
 45 35 40 45
 Ala Lys Ser His His Asp Leu Leu Thr Arg His Pro Tyr Glu Thr Ala
 50 55 60
 Ala Pro Leu Leu Ser Ser His Leu Ile Leu Thr Glu Ala Leu Arg Asn
 65 70 75 80
 50 Gly Leu Gly Lys Cys His Asp Pro His Phe Thr Gly Glu Glu Thr Glu
 85 90 95
 Ala Gln Arg Gly Lys Leu Thr Thr
 100
 55 <210> 479
 <211> 439
 <212> PRT
 <213> Homo sapiens
 60 <400> 479
 Leu Gly Asp His Gly Trp Glu Leu Ser Leu Glu Glu Asp Ala Gln Leu
 1 5 10 15
 Trp Gly Gly Val Val Lys Ser Cys Phe Glu Gly Lys Gly Pro Gln Arg

20 25 30
 Glu Ala Gln Pro Ala Ser Pro Gln Ala Ala Pro Pro Gly Pro Thr Asn
 35 40 45
 5 Glu Ala Gln Met Ala Ala Ala Ala Leu Ala Arg Leu Glu Gln Lys
 50 55 60
 Gln Ser Arg Ala Trp Gly Pro Thr Ser Gln Asp Thr Ile Arg Asn Gln
 65 70 75 80
 Val Arg Lys Glu Leu Gln Ala Glu Ala Thr Val Ser Gly Ser Pro Glu
 85 90 95
 10 Ala Pro Gly Thr Asn Val Val Ser Glu Pro Arg Glu Glu Gly Ser Ala
 100 105 110
 His Leu Ala Val Pro Gly Val Tyr Phe Thr Cys Pro Leu Thr Gly Ala
 115 120 125
 Thr Leu Arg Lys Asp Gln Arg Asp Ala Cys Ile Lys Glu Ala Ile Leu
 130 135 140
 15 Leu His Phe Ser Thr Asp Pro Val Ala Ala Ser Ile Met Lys Ile Tyr
 145 150 155 160
 Thr Phe Asn Lys Asp Gln Asp Arg Val Lys Leu Gly Val Asp Thr Ile
 165 170 175
 20 Ala Lys Tyr Leu Asp Asn Ile His Leu His Pro Glu Glu Glu Lys Tyr
 180 185 190
 Arg Lys Ile Lys Leu Gln Asn Lys Val Phe Gln Glu Arg Ile Asn Cys
 195 200 205
 25 Leu Glu Gly Thr His Glu Phe Phe Glu Ala Ile Gly Phe Gln Lys Val
 210 215 220
 Leu Leu Pro Ala Gln Asp Gln Glu Asp Pro Glu Glu Phe Tyr Val Leu
 225 230 235 240
 Ser Glu Thr Thr Leu Ala Gln Pro Gln Ser Leu Glu Arg His Lys Glu
 245 250 255
 30 Gln Leu Leu Ala Ala Glu Pro Val Arg Ala Lys Leu Asp Arg Gln Arg
 260 265 270
 Arg Val Phe Gln Pro Ser Pro Leu Ala Ser Gln Phe Glu Leu Pro Gly
 275 280 285
 Asp Phe Phe Asn Leu Thr Ala Glu Glu Ile Lys Arg Glu Gln Arg Leu
 290 295 300
 35 Arg Ser Glu Ala Val Glu Arg Leu Ser Val Leu Arg Thr Lys Ala Met
 305 310 315 320
 Arg Glu Lys Glu Glu Gln Arg Gly Leu Arg Lys Tyr Asn Tyr Thr Leu
 325 330 335
 40 Leu Arg Val Arg Leu Pro Asp Gly Cys Leu Leu Gln Gly Thr Phe Tyr
 340 345 350
 Ala Arg Glu Arg Leu Gly Ala Val Tyr Gly Phe Val Arg Glu Ala Leu
 355 360 365
 Gln Ser Asp Trp Leu Pro Phe Glu Leu Leu Ala Ser Gly Gly Gln Lys
 370 375 380
 45 Leu Ser Glu Asp Glu Asn Leu Ala Leu Asn Glu Cys Gly Leu Val Pro
 385 390 395 400
 Ser Ala Leu Leu Thr Phe Ser Trp Asp Met Ala Val Leu Glu Asp Ile
 405 410 415
 50 Lys Ala Ala Gly Ala Glu Pro Asp Ser Ile Leu Lys Pro Glu Leu Leu
 420 425 430
 Ser Ala Ile Glu Lys Leu Leu
 435
 55 <210> 480
 <211> 116
 <212> PRT
 <213> Homo sapiens
 60 <400> 480
 Met Trp Ala Arg Leu Pro His Thr Pro Glu Gln Met Gly His Arg Leu
 1 5 10 15
 Ile Gly Pro Lys Glu Ala Ser Leu His Val Val Pro Ser Trp Pro Ala

20 25 30
 Arg Lys Met Glu Gly Leu Leu Ala Gly Leu Ser Ser Ser Pro Arg Lys
 35 40 45
 Ser Cys Trp Pro Phe Trp Val His Gly Pro Lys Val His Glu Gly Gly
 5 50 55 60
 Ser Ala Cys Glu Thr Ser Ser Ser Trp Val Glu Gly Leu Gly Leu Arg
 65 70 75 80
 Arg Val Thr Ser Val His Ser Leu Cys Gln Gly Leu Gly Ala Ser Val
 85 90 95
 10 Gln Leu Leu Pro Gly Pro Pro Pro Thr Thr Thr Ser Asp Lys Asn Asn
 100 105 110
 Tyr Thr Ser Gly
 115

15 <210> 481
 <211> 171
 <212> PRT
 <213> Homo sapiens

20 <400> 481
 Met Gln Pro Ala Glu Arg Ser Arg Val Pro Arg Ile Asp Pro Tyr Gly
 1 5 10 15
 Phe Glu Arg Pro Glu Asp Phe Asp Asp Ala Ala Tyr Glu Lys Phe Phe
 20 25 30
 25 Ser Ser Tyr Leu Val Thr Leu Thr Arg Arg Ala Ile Lys Trp Ser Arg
 35 40 45
 Leu Leu Gln Gly Gly Gly Val Pro Arg Ser Arg Thr Val Lys Arg Tyr
 50 55 60
 Val Arg Lys Gly Val Pro Leu Glu His Arg Ala Arg Val Trp Met Val
 30 65 70 75 80
 Leu Ser Gly Ala Gln Ala Gln Met Asp Gln Asn Pro Gly Tyr Tyr His
 85 90 95
 Gln Leu Leu Gln Gly Glu Arg Asn Pro Arg Leu Glu Asp Ala Ile Arg
 100 105 110
 35 Thr Asp Leu Asn Arg Thr Phe Pro Asp Asn Val Lys Phe Arg Lys Thr
 115 120 125
 Thr Asp Pro Cys Leu Gln Arg Thr Leu Tyr Asn Val Leu Leu Ala Tyr
 130 135 140
 Gly His His Asn Gln Gly Val Gly Tyr Cys Gln Gly Met Asn Phe Ile
 40 145 150 155 160
 Ala Gly Tyr Leu Ile Leu Ile Thr Asn Asn Glu
 165 170

<210> 482
 45 <211> 177
 <212> PRT
 <213> Homo sapiens

<400> 482
 50 Met Gln Pro Ala Glu Arg Ser Arg Val Pro Arg Ile Asp Pro Tyr Gly
 1 5 10 15
 Phe Glu Arg Pro Glu Asp Phe Asp Asp Ala Ala Tyr Glu Lys Phe Phe
 20 25 30
 Ser Ser Tyr Leu Val Thr Leu Thr Arg Arg Ala Ile Lys Trp Ser Arg
 55 35 40 45
 Leu Leu Gln Gly Gly Gly Val Pro Arg Ser Arg Thr Val Lys Arg Tyr
 50 55 60
 Val Arg Lys Gly Val Pro Leu Glu His Arg Ala Arg Val Trp Met Val
 65 70 75 80
 60 Leu Ser Gly Ala Gln Ala Gln Met Asp Gln Asn Pro Gly Tyr Tyr His
 85 90 95
 Gln Leu Leu Gln Gly Glu Arg Asn Pro Arg Leu Glu Asp Ala Ile Arg
 100 105 110

	Thr	Asp	Leu	Asn	Arg	Thr	Phe	Pro	Asp	Asn	Val	Lys	Phe	Arg	Lys	Thr
			115					120					125			
	Thr	Asp	Pro	Cys	Leu	Gln	Arg	Thr	Leu	Tyr	Asn	Val	Leu	Leu	Ala	Tyr
			130				135					140				
5	Gly	His	His	Asn	Gln	Gly	Val	Gly	Tyr	Cys	Gln	Gly	Met	Asn	Phe	Ile
	145					150					155					160
	Ala	Gly	Tyr	Leu	Ile	Leu	Ile	Thr	Asn	Asn	Asp	Lys	Asn	Leu	Phe	Gly
					165					170					175	
	Cys															

THIS PAGE BLANK (USPTO)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
14 June 2001 (14.06.2001)

PCT

(10) International Publication Number
WO 01/42451 A3

- (51) International Patent Classification⁷: C12N 15/09, (74) Common Representative: GENSET; Intellectual Property Department, 24, rue Royale, F-75008 Paris (FR).
C07K 14/47
- (21) International Application Number: PCT/IB00/01938 (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (22) International Filing Date: 7 December 2000 (07.12.2000)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/169,629 8 December 1999 (08.12.1999) US
60/187,470 6 March 2000 (06.03.2000) US
- (71) Applicant (*for all designated States except US*): GENSET [FR/FR]; Intellectual Property Department, 24, rue Royale, F-75008 Paris (FR).
- (72) Inventors; and
- (75) Inventors/Applicants (*for US only*): DUMAS MILNE EDWARDS, Jean-Baptiste [FR/FR]; 8, rue Grégoire de Tours, F-75006 Paris (FR). BOUGUELERET, Lydie [FR/FR]; 108, avenue Victor Hugo, F-92170 Vanves (FR). JOBERT, Séverin [FR/FR]; 7, impasse Tourneux, F-75010 Paris (FR).
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW). Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM). European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR). OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- Published:
— with international search report
- (88) Date of publication of the international search report:
10 May 2002
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*



WO 01/42451 A3

(54) Title: FULL-LENGTH HUMAN cDNAs ENCODING POTENTIALLY SECRETED PROTEINS

(57) Abstract: The invention concerns GENSET polynucleotides and polypeptides. Such GENSET products may be used as reagents in forensic analyses, as chromosome markers, as tissue/cell/organelle-specific markers, in the production of expression vectors. In addition, they may be used in screening and diagnosis assays for abnormal GENSET expression and/or biological activity and for screening compounds that may be used in the treatment of GENSET-related disorders.

INTERNATIONAL SEARCH REPORT

International Application No

PCT/IB 00/01938

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/09 C07K14/47

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, PAJ, BIOSIS, MEDLINE, EMBL

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	DATABASE GENESEQ [Online] AC No. Y76144, 23 March 2000 (2000-03-23) RUBEN SM ET AL: "Human secreted protein encoded by gene 21" XP002163702 polypeptide 100% identical with SEQ ID No:242 over its entire length (508 aa) abstract ---	1-26
P,X	DATABASE GENESEQ [Online] AC No. Z65270, 23 March 2000 (2000-03-23) RUBEN SM ET AL: "Human secreted protein gene 21" XP002163703 Polynucleotide with 99.9% identity to SEQ ID No:1 over 2138 nucleotides abstract --- -/--	1-26



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

25 May 2001

Date of mailing of the international search report

30.05.01

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040. Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Nichogiannopoulou, A

INTERNATIONAL SEARCH REPORT

International Application No

PCT/IB 00/01938

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>DATABASE EMBL [Online] AC No. AI911546, 30 July 1999 (1999-07-30) NATL. CANCER INST.: "ty73d05.x1 NCI CGAP Kid11 Homo sapiens cDNA clone IMAGE:2284713" XP002163704 Polynucleotide 100% identical with SEQ ID No:1 over 519 nucleotides abstract</p>	1-10, 15-17
X	<p>--- DATABASE EMBL [Online] AC No. AI361251, 7 January 1999 (1999-01-07) NATL. CANCER INST.: "qy42e02.x1 NCI CGAP Brn23 Homo sapiens cDNA clone IMAGE:2014682" XP002163705 Polynucleotide 100% identical to SEQ ID No:1 over 454 nucleotides abstract</p>	1-10, 15-17
A	<p>--- WO 97 07198 A (GENETICS INST) 27 February 1997 (1997-02-27) the whole document</p>	1-26
A	<p>--- JACOBS K A ET AL: "A GENETIC SELECTION FOR ISOLATING CDNAS ENCODING SECRETED PROTEINS" GENE,NL,ELSEVIER BIOMEDICAL PRESS. AMSTERDAM, vol. 198, 1 October 1987 (1987-10-01), pages 289-296, XP002045919 ISSN: 0378-1119 the whole document</p>	1-26
A	<p>--- TASHIRO K ET AL: "SIGNAL SEQUENCE TRAP: A CLONING STRATEGY FOR SECRETED PROTEINS AND TYPE I MEMBRANE PROTEINS" SCIENCE,US,AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE,, vol. 261, 30 July 1993 (1993-07-30), pages 600-603, XP000673204 ISSN: 0036-8075 the whole document</p> <p>-----</p>	1-26

INTERNATIONAL SEARCH REPORT

Int. application No.
PCT/IB 00/01938

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

B x II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-26 all partially

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

Invention 1: Claims: 1-26 (all partially)

Isolated polynucleotides, polypeptides, expression vectors, host cells, non-human transgenic animals methods and antibodies relating to the nucleotide sequence of SEQ ID No:1 and the amino acid sequence of SEQ ID No:242

Inventions 2 to 241: Claims: 1-26 (all partially)

Idem as invention 1, but relating to each of the nucleotide sequences of SEQ ID Nos:2 to 241 and the corresponding amino sequences of SEQ ID Nos:243 to 482, where invention 2 is limited to SEQ ID Nos:2 and 243, invention 3 is limited to SEQ ID Nos:3 and 244,....., and invention 241 is limited to SEQ ID Nos:241 and 482.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/IB 00/01938

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9707198 A	27-02-1997	US 5707829 A	13-01-1998
		AU 727480 B	14-12-2000
		AU 6712396 A	18-02-1997
		AU 727489 B	14-12-2000
		AU 6768596 A	12-03-1997
		CA 2227220 A	06-02-1997
		CA 2229208 A	27-02-1997
		EP 0839196 A	06-05-1998
		EP 0851875 A	08-07-1998
		JP 11510045 T	07-09-1999
		US 6043344 A	28-03-2000
		WO 9704097 A	06-02-1997
		US 6074849 A	13-06-2000
		US 5969093 A	19-10-1999